2019

Structural Investigation of BcsC: Insight into Periplasmic Transport During Cellulose Export

William Scott
scot0830@mylaurier.ca

Joel T. Weadge
Wilfrid Laurier University

Structural Investigation of BcsC: Insight into Periplasmic Transport During Cellulose Export

by

William G. J. Scott

Honors BSc Biology & Psychology,

Wilfrid Laurier University, 2015

Thesis

Submitted to the Faculty of Biology

in partial fulfillment of the requirements for

Master of Science Integrative Biology,

Wilfrid Laurier University

Acknowledgements

I would like to primarily thank Dr. Joel Weadge for giving me the opportunity to pursue research and for his painstaking efforts in pushing me to grow and succeed during this degree. I would also like to thank the individuals on my committee Dr. Robin Slawson and Dr. Mike Suits for their time, effort, and help throughout this process. Also, I would like to thank all the members of the Weadge lab for all their assistance throughout the course of this degree as well as all the members of the other labs involved in research on the 4th floor, namely the Suits lab, the Slawson lab, and the Smith lab as well as the Horseman lab on the 1st floor. In addition, I would like to thank my family and friends for having patience and supporting me during this time-consuming endeavour.

Abstract

A biofilm can be defined by a community of microbes coexisting within a self-produced protective polymeric matrix. Exopolysaccharide (EPS) is a key component in biofilms and a contributor to their virulence and pathogenicity. The cellulose bacterial synthesis complex is one such EPS system that is found in many *Enterobacteriaceae*, including *Escherichia coli* and *Salmonella* spp., and is responsible for the production and secretion of the EPS cellulose. BcsC is the periplasmic protein responsible for the export of the exopolysaccharide cellulose and was the focus of this research. Sequence homology comparisons and structural predictions between BcsC, and the previously characterized alginate export proteins AlgK and AlgE indicate similar roles in facilitating the translocation of EPS across the bacterial cell wall. However, there are discrepancies between the systems, such as the purpose of several additional tetratricopeptide regions (TPRs) contained within BcsC compared to AlgK. To better understand the role that BcsC plays in cellulose export structural characterization of this protein was pursued. Six protein constructs that together cover overlapping portions of BcsCs TPR region were successfully expressed and purified, four of which were further analyzed with SAXS and screened for crystal formation. SAXS data was merged with a pre-existing protein data bank file of $BcsC^{TPR\ 1\text{-}6}$ to identify similar regions and provided conceptual renderings as to the orientation and size of the protein. Promising crystal hits from $BcsC^{TPR\ 12\text{-}21}$ and $BcsC^{TPR\ 1\text{-}15}$ were obtained, optimized and sent for X-ray diffraction, with resolution results between 12 and 2.8 Å. A complete dataset for $BcsC^{TPR\ 1\text{-}15}$ has since been collected and structure solution is ongoing through a combination of molecular replacement and selenomethionine (SeMet) labelling techniques. Preliminary SeMet crystals are promising, but currently appear thinner than native crystals and additional optimization may be required before suitable X-ray diffraction data can be obtained.

List of Abbreviations

A_{280}: Absorbance at 280nm wavelength

BcsC: Bacterial cellulose synthase protein C

CBDs: Carbohydrate Binding Domains

c-di-GMP: Bis-(3´-5´)-cyclic dimeric guanosine monophosphate

CF: Cystic Fibrosis

C-terminal: Carboxyl terminal

DMSO: Dimethyl sulfoxide

DLS: Dynamic light scattering

EPS: Exopolysaccharide

FPLC: Fast protein liquid chromatography

His_6-tag: Hexahistidine affinity tag

IMAC: Immobilized metal ion affinity chromatography

IPTG: Isopropyl β-D-1-thiogalactopyranoside

kDa: Kilodalton

Kpsi: Kilopounds per square inch

LB: Luria-Bertani media

Ni-NTA: Nickel-nitrilotriacetic acid chelate

N-terminal: Amino-terminal

OD_{600}: Optical density at 600nm wavelength

PDB: Protein data bank

Phyre2: Protein homology/analogy recognition engine, version 2

pI: Isoelectric point

SAXS: Small angle X-ray scattering

SB: Super broth media

SDS-PAGE: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SeMet: Selenomethionine

TPR: Tetratricopeptide repeat [motif]

Tris: Tris(hydroxymethyl)aminomethane

UDP: Uridine diphosphate

Table of Contents

List of Tables

List of Figures

1.0 Introduction

1.1 *Escherichia coli* and *Salmonella* spp.

Escherichia coli and *Salmonella* spp. are among the leading causes of foodborne diseases. Foodborne diseases are estimated to cause illness in as many as 600 million people, of which approximately 420,000 people die annually, with almost one third of them being children under five years of age (1). The United States alone has an estimated 48 million cases annually of foodborne illness, with a predicted economic impact of 77.7 billion dollars in health-related costs (2). In the United States, *Salmonella* accounts for approximately 3.6 billion dollars per year of these health-related costs (3). Without intervention the statistics may only increase, which makes research towards the goal of alleviating the economic burden and physical pressures related to this crisis of paramount importance.

While *Salmonella* spp. are typically viewed as pathogenic (4), many strains of *E. coli* are harmless and are part of the natural microbiota of the mammalian intestinal tract (5). However, some strains of *E. coli* can be deadly to human hosts (*ie.*, strain o157:H7;) and have been responsible for numerous outbreaks of contaminated food and water illnesses (6). A local outbreak was exemplified by the small rural town of Walkerton, Ontario, where a pathogenic strain of *E. coli* led to an epidemic causing the death of 7 individuals and 2,300 others became ill through water consumption (7). This strain of *E. coli* (o157:57) was alone responsible for 390 outbreaks in the United States between 2003 and 2012 (8). More recent outbreaks of *E. coli* infections have been linked to ground beef (9) and two separate flour product recalls, one in the U.S (9) and in Canada (10). Evidence of outbreaks from multidrug-resistant strains of *Salmonella* (11) call further attention to the need for understanding the mechanisms by which

bacteria survive in a host during an infection, but also the mechanism by which these organisms persist in our food/water distribution systems is necessary.

1.2 Biofilms

Bacteria exist in a unicellular state where the planktonic cells are free swimming and can form a multicellular complex, where the cells are sessile and exist in a biofilm (12). Bacterial biofilms can be described as heterogeneous structures that may contain different populations of microorganisms, which are encompassed by a matrix (composed primarily of exopolysaccharides) that facilitates attachment to a variety of surfaces and interfaces (including both inert and/or organic) to facilitate growth as a micro-ecosystem where by-products from different organisms are cross-utilized (13, 14). Biofilm formation is the primary survival strategy utilized by the majority of infectious bacteria (15–17). Biofilms are prominent within the human body persisting as chronic infections, entrenched on medical devices, and in the oral cavity (15). In these environments, biofilms contribute resistance to disinfectants (like chlorine) of up to 1000 times (14). According to Lewis (2001), an estimated 60% of infections in humans are due to the formation of biofilms; thus increasing the necessity for research and strategies in prevention and treatment (19). Biofilms also permit bacteria to flourish in a wide variety of environments, where pH, temperature, humidity, nutrient content and other factors vary widely (20–22). These protective features of a biofilm are particularly important from a health perspective, as the biofilm matrix also increases bacterial resistance to immune detection/clearance, antibiotics and disinfectants (15, 17, 21, 23).

1.3 Biofilm Formation and the role of c-di-GMP

The second messenger c-di-GMP is a widely conserved soluble intracellular signaling molecule of paramount importance in controlling biofilm formation (21). The regulatory

functions of c-di-GMP production are controlled by several effectors at multiple levels (transcriptional, translational, and at the protein level). For example, two groups of enzymes with antagonistic activities exert control over the c-di-GMP turnover rate in the bacterial cytoplasm (25, 26). A c-di-GMP molecule is enzymatically synthesized from two guanosine triphosphate (GTP) molecules through the action of diguanylate cyclases (DGC), which contain a GGDEF amino acid consensus sequence domain; whereas, the hydrolysis of c-di-GMP into linear pGpG or two guanosine monophosphate (GMP) molecules is controlled by phosphodiesterases (PDE) that contain EAL or HD-GYP domains (27–31). The cytoplasmic flux in concentrations of c-di-GMP oscillates with environmental changes exerted by the activites of these opposing enzymes (DGC and PDE proteins) and is directly related to the production of exopolysaccharides (EPSs), virulence, adhesion, motility, and cell morphogenesis in a diverse range of bacterial species including *E. coli* and *Salmonella* spp. (24, 26, 30, 32, 33). C-di-GMP, through binding to riboswitches and acting as an effector to a variety of proteins, is responsible for regulating multiple two component systems (34–36). For example, high concentrations of intracellular c-di-GMP will typically lead to EPS synthesis, loss of motility (flagella), increased adhesion and aggregation; thereby favouring biofilm formation (24, 26, 32); whereas, a low concentration results in the opposite effect, motility and highly invasive phenotypes. Therefore, given that c-di-GMP is highly conserved across a diverse range of bacterial species, this signaling molecule represents a common mechanism of regulation and control over biofilm formation and dispersal (32, 37).

Once c-di-GMP levels rise, biofilm formation by independent cells is triggered to begin, and colony growth progresses through four main stages of development (38, 39). The main stages of biofilm formation are attachment, aggregation, maturation, and dispersal (Figure 1).

Attachment of a cell to a substratum, also known as adhesion, is mediated by the characteristics of the surface and the cell (13, 40). The properties of the surface that might affect attachment are hydrophobicity, charge, and texture (22), whereas, the characteristics of the cell surface that would impact attachment are the presence of pili, fimbriae, flagella and/or extracellular polymeric substances (41, 42). These bacterial macromolecules are important factors in promoting attachment by aiding the bacteria in overcoming repulsive forces or binding specific structures on a surface until permanent attachment features are in place/expressed (18). Notably, the initial step is called reversible attachment (Figure 1; panel A), while the second step of biofilm formation, known as aggregation, is where irreversible attachment takes place.

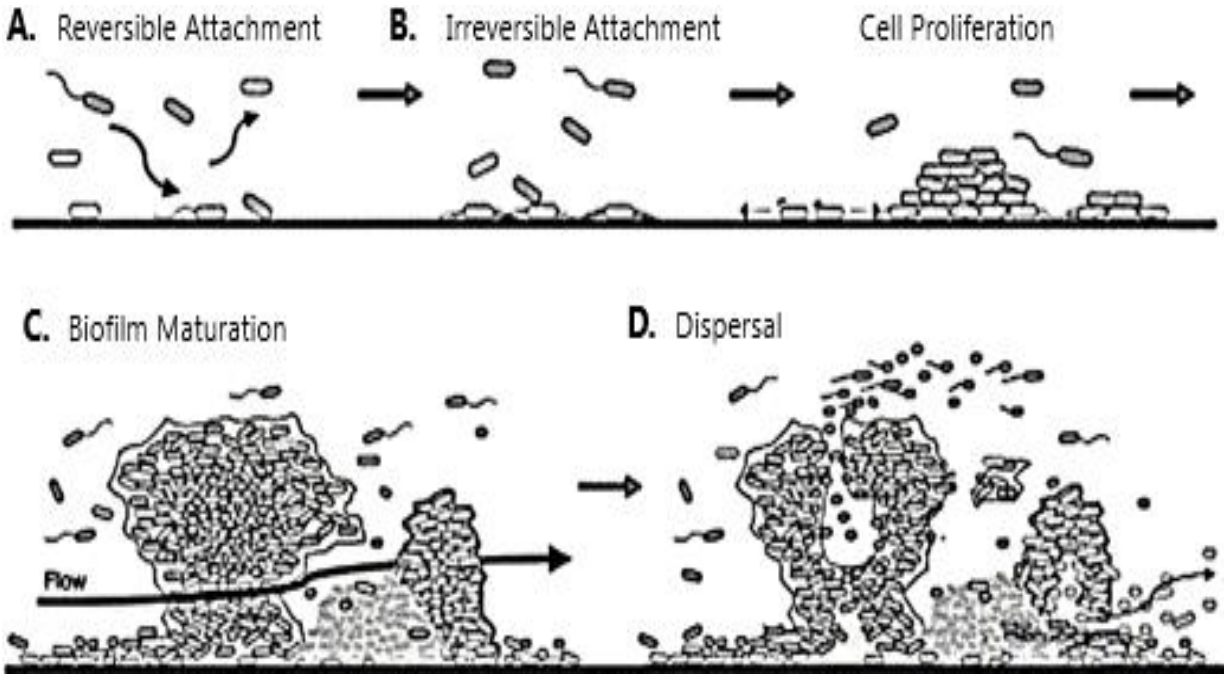


Figure 1. Stages of Biofilm Formation. Biofilm development is initiated by (A) reversible attachment of individual cells to the substratum, which may require motility. The second step (B) in attachment is irreversible adherence with exopolysaccharide and pili and loss of motility appendages, followed by aggregation through cell proliferation. The third stage (C) is marked by the continued growth of the biofilm, presence of solvent channels and heterogenic biofilm structures that become established along with production of extracellular polymeric matrix and cell-cell interactions. The final stage (D) illustrates individual bacteria being released from the biofilm that may again begin the process somewhere else (modified from (43)).

The second stage of biofilm formation is signified by the alteration of the cell surface that assists with the development of a monolayer and initial aggregation of the bacteria (Figure 1; Panel B) (41). For example, intracellular increase of c-di-GMP levels influence the downregulation of motility appendages and the upregulation of EPS and fimbriae/pili; thereby leading to irreversible attachment to a surface (38, 44, 45). Bacteria then continue to aggregate and adhere to one another resulting in the creation of microcolonies (13). The maturation stage of the biofilm matrix leads to the continued growth of microcolonies into multilayered structures where solvent channels begin to form and are actively maintained to allow bacteria access to nutrients, water, and oxygen (Figure 1; Panel C) (15, 46). As the biofilm continues to mature, many different shapes can be adopted that are often influenced by environmental factors, such as pH, ionic strength, hydrodynamic shear, nutrient supply, species present in the biofilm, surface temperature, and/or host conditions (13, 20–22). For example, specific conditions lead to the formation of monolayers, multilayers (mushroom or pillar like 3D structures), and even pellicles at the air-liquid interface (23, 47).

In mature biofilms, particularly within types that are involved in chronic infections, there exists a subtype of persister cells (48). These cells are not mutants, but instead are phenotypic variants of wildtype cells (48, 49). Persister cells are believed to be created by an accumulation of toxins in the bacterial cell that forces the cell into metabolic stasis (49, 50). While dormant, these cells may encounter antibiotics, yet often are not destroyed because the cell is not actively metabolizing or dividing. Persister cells can, thus, be multidrug tolerant without acquiring any additional defence mechanisms for subverting antibiotics (50). In addition, mature biofilms are known for fostering conditions ideal for horizontal gene transfer (conjugation, transfection and/or transformation) in which antimicrobial resistant genes can be passed rapidly throughout

members of the biofilm (23, 43). Thus, metabolizing cells can also actively acquire antimicrobial resistance in a biofilm setting. The rise of antimicrobial resistance in conjunction with biofilm formation has led to an increasing concern from a public health view.

As the mature biofilm grows in size, some cells are separated from nutrient sources or exposed to detrimental environmental conditions and, thus, the bacteria have developed mechanisms to disperse from the biofilm (Figure 1; Panel D) (39, 47). In this final stage of biofilm development, termed dispersal, the bacteria leave the biofilm through desorption/detachment (passive separation) or through dispersion (active separation) (39). Active dispersal mechanisms are initiated by the bacteria themselves (*eg.*, enzymatic or chemical), while passive separation (*eg.*, erosion or sloughing) are due to external forces (51). As the bacteria make the transition from sessile to free swimming, they increase the expression of certain appendages for propulsion, and downregulate the production of EPS and other irreversible attachment structures (38, 47, 52).

1.4 Bacterial EPS

P. aeruginosa is one of the most widely studied organisms partly due to the role of the bacterial biofilm produced by this organism during infection of Cystic Fibrosis (CF) patients (53). In the CF lung, a human genetic mutation within the CF transmembrane conductance regulator (CFTR) gene results in improper clearance of mucus secretions from airway passages (53, 54). Accumulation of mucus provides an ideal environment for the attachment and colonization of opportunistic organisms, like *P. aeruginosa* (54). Within the CF lung, *P. aeruginosa* exacerbates the problem due to the production of a biofilm, which is predominantly composed of the EPS, alginate (composed of mannuronic and guluronic acid) (54). Production of alginate during infection of the CF lung is often linked to a poor prognosis for the patient (55).

As a result, there has been a wealth of research surrounding the production and effects of this polymer in conjunction with improving mitigation strategies for infection by *P. aeruginosa*.

Bacterial cellulose is also an EPS that has numerous applications and economical significance. Cellulose, characterized by β(1-4)-glycosidic linkages, is the most abundant biopolymer on earth, and is found in vascular plants, algae, and bacteria (56–58). Plant cellulose has the additional polymers, lignins and hemicelluloses present, whereas, bacterial cellulose is formed as a pure polymer (59). Bacterial cellulose contains hydrogen bonds between the fibrillar units that provide strength and flexibility, allowing changes in shape to conform to different surfaces (60). Bacterial cellulose is naturally hydrophilic and is ordered into nano and microfibrils, which can allow for the formation of a hydrogel due to the binding of large amounts of water (59, 60). These examples only hint at the dozens of potential applications in drug design and delivery systems, cosmetics, as well as food and food packaging materials that have been proposed to be possible with this polymer (61). Furthermore, bacterial cellulose has promising biomedical applications due to its low immunogenic potential and biocompatibility (62).

In addition to its economic importance, bacterial cellulose has serious health impacts and has been identified in many bacteria including several species of *Gluconacetobacter xylinus* (63), *Enterobacter* (64, 65), *Salmonella* spp. (66), and *E. coli.* (5, 66). Although not all bacteria that produce cellulose biofilms are harmful, several species of *Salmonella* spp. have proven to be virulent within a host. For example, a study by Solano and colleagues (2001), found that biofilm-forming strains of *Salmonella enteritidis* were highly virulent compared to non-biofilm forming strains when tested in a chicken infection model (67). A subsequent study by Solano and colleagues (2002), indicated that cellulose was not required for virulence in *S. enteritidis* (68). Using a chicken infection model, cellulose deficient strains were still virulent, yet the negative

cellulose mutant strains were highly susceptible to chlorine treatment where the wildtype was chlorine resistant (68). Taken together, these studies suggest that although cellulose production may not be required for virulence, it will increase survival. Moreover, Domenico and colleagues (2017) have suggested a multistage strategy for *Salmonella typhi:* colonization, followed by chronic persistence, and toxin production (4). *S. typhi* is known for its role in gallbladder cancer due to release of a carcinogenic toxin and has the capacity for survival as a biofilm within the gallbladder and on gallstones (4). Biofilms have been found in asymptomatic and symptomatic carriers in the harsh high-bile environment with antimicrobial properties, which is in a location that provides a direct route to release toxins to several susceptible target areas (4). Thus, as the biofilm provides the survivability to invade and colonize the host, the bacteria can then detach from biofilm and invade host cells while other the bacteria maintain the biofilm, leading to the process of chronic persistence (4). From the literature, many different applications for bacterial cellulose exist, in addition to the evidence that cellulose plays an important part in the chronic persistence of biofilms that can maintain virulence, thereby, making the understanding of its synthesis and export from the bacterial cell of consequential significance.

1.5 EPS biosynthesis

The focus of the present research is the cellulose biosynthesis pathway. However, many of the biosynthetic steps of microbial cellulose have been inferred from studies on the more characterized alginate biosynthetic pathway (69). The inference between the two pathways is due to the similarity between the proteins and enzymatic steps that occur in the synthesis and export of the polymers across the bacterial cell wall (Figure 2). Thus, a brief analysis of the alginate

biosynthetic pathway will first be presented, followed by the cellulose biosynthesis pathway.

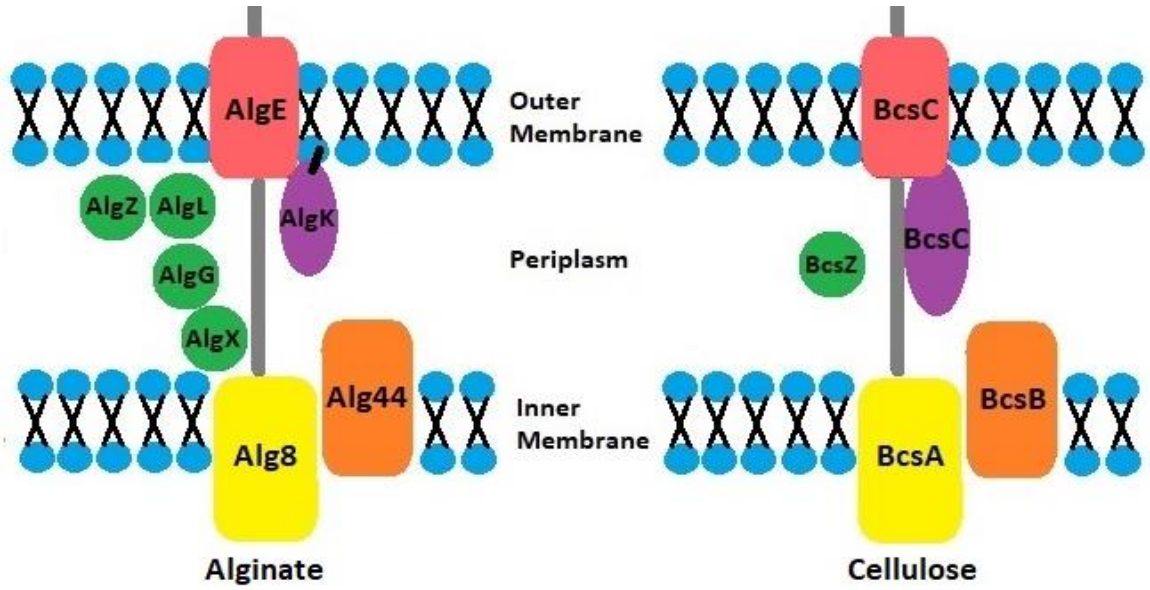


Figure 2. Graphic Representation of the Alginate and Cellulose Biosynthetic Systems. Each pathway component is indicated on the schematic and colour-coded according to predicted similar functions as follows: green, exopolysaccharide modification enzymes; yellow, synthase; orange, transmembrane anchor; red, β-barrel porin; purple, TPR; The black line specifies the N-terminal lipid anchor of AlgK. Please note that BcsC is one continuous protein but it was colour coded with the alginate system to show homologous regions. Through the two systems, the polysaccharide indicated is polymerized and transported via its respective synthase following c-di-GMP binding. In the periplasm, polysaccharide modifying enzymes act on the polysaccharides prior to export by the TPR/ β-barrel protein regions (modified from (69)).

The *algD* operon (*alg824KEGXLIJFA*) encodes the proteins required for the polymerization and export of alginate (69). Briefly, poly-β-D-mannuronic acid is synthesized from guanosine 5'-diphosphate-mannuronic acid at the inner membrane, while its polymerization and export is believed to be facilitated by the synthetase Alg8 and the c-di-GMP receptor Alg44 (70, 71). Polymerization and transport across the inner membrane is thought to be regulated by the binding of c-di-GMP to the PilZ domain on Alg44 (70). Additional evidence from deletion mutants suggest that AlgX and AlgG may assist in directing the alginate polymer across the periplasm, as well as protecting it from the periplasmic alginate lyase, AlgL (72, 73). The outer-

membrane lipoprotein AlgK (Figure 3), thought to be composed of 10 TPR domains, is proposed to act as a scaffold where other periplasmic Alg proteins can interact (74). AlgK is proposed to conduct the mature alginate polymer to the integral outer-membrane 18-stranded β-barrel protein, AlgE, which enables the passage of alginate through the outer membrane (75). Although a lack of definitive evidence of a direct interaction between AlgK and AlgE exists, support from mutant and localization studies suggest that AlgK contributes to the localization of AlgE (69). Additionally, evidence from homologous proteins, BcsC, PelB, and PgaA (involved in the export of cellulose, Pel polysaccharide, and poly β-1-6-GlcNAc, respectively), are all predicted through bioinformatics analyses to be large two-domain proteins with tandem-TPR and β-barrel regions, analogous to a fusion of the AlgK/AlgE protein complex (69).

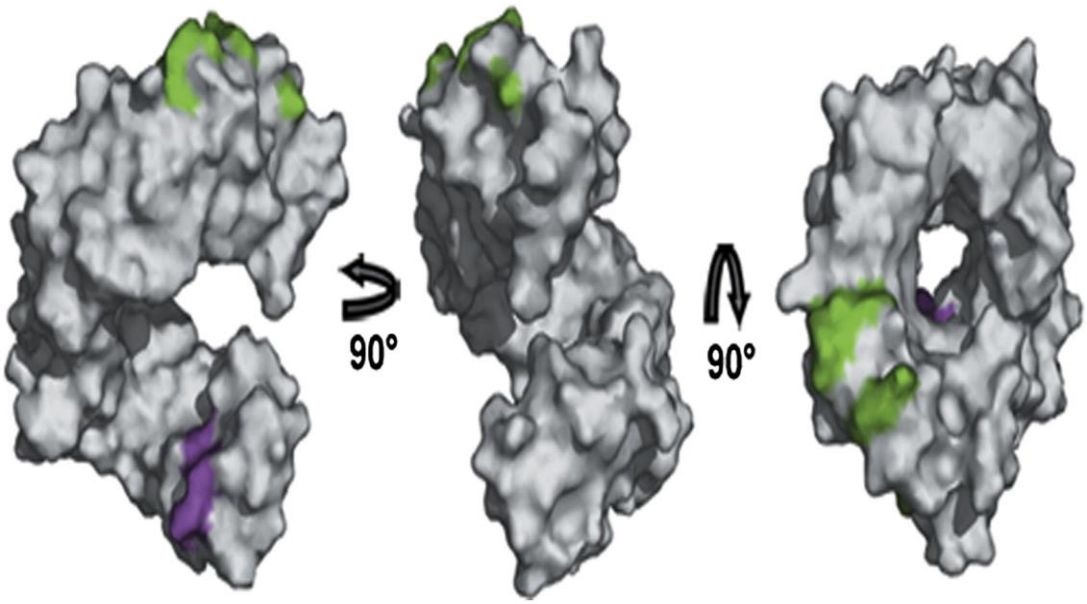


Figure 3. A Surface Model Illustration of AlgK in Three Orthogonal Alignments. Conserved residues on the surface of the protein have been identified by the researchers as possible binding sites for protein-protein interactions and are indicated with purple and green colour (modified from (74)).

Cellulose synthesis is encoded by the bacterial cellulose synthesis (bcs) operon *bcsABZC*, of which BcsA and BcsB are known as the catalytic core of the inner membrane (analogous to Alg8/44 from the alginate system – Figure 2) (76). BcsA contains a PilZ domain at its C-terminus that binds c-di-GMP, and catalyzes cellulose polymerization from UDP-activated glucose in a process similar to that of the alginate system (31, 70, 77, 78). BcsB, a periplasmic protein anchored to the inner membrane by a transmembrane helix, provides stability to the transmembrane region of BcsA that is essential for catalysis (77). BcsB is theorized to initially direct the cellulose polymer from the cytoplasmic membrane into the periplasm through two carbohydrate-binding domains (77). BcsZ is a periplasmic glycosyl-hydrolase with endo-β-1,4-glucanase activity that is believed to facilitate the degradation of EPS much like AlgL (72, 79), yet little is known regarding its interaction with BcsB-BcsA or BcsC (74, 76, 78). In the periplasm, BcsC has a TPR domain analogous to AlgK (Figure 3) that is likely responsible for mediating the passage of EPS through the periplasm to the 16 stranded β-barrel region (analogous to that of AlgE) that traverses the outer membrane (75, 78). Of particular note to this thesis is that homologous proteins from separate EPS biosynthetic systems (*eg.*, cellulose, Pel polysaccharide, and poly β-1-6-GlcNAc) have conserved TPR regions of varying lengths that may correspond to specific requirements for each of these systems. Indeed, a review of the literature on TPRs (outlined below) also indicates that TPRs can serve multiple roles through the binding of protein, carbohydrates and even DNA, but these possibilities are still relatively unknown for the related proteins (BcsC, PelA and AlgK) from the EPS biosynthetic systems.

1.6 Tetra-trico-peptide repeats (TPRs)

Tetra-trico-peptide repeats (TPRs) are present in a wide variety of proteins involved in many different functions (80–83). This is due to the capacity to serve as a protein scaffold and

mediate protein-protein and protein-carbohydrate interactions (80–82, 84). A TPR is a versatile structural motif that consists of 34 amino acid tandem repeats (80, 85) comprising a degenerate consensus sequence of small and large hydrophobic amino acids (Figure 4; 82, 83). Interestingly, there are no completely invariant residues present, yet some residues are highly conserved in TPR domains (83, 86). The tertiary structure of the canonical TPR motifs is a helix-turn-helix fold (83). The adjacent parallel packing of TPR units adopt a succession of repeating anti-parallel α-helices and produces an overall super-helical structure (83). The residue between adjacent TPR motifs affects the type of twist given to the super-helix (82, 83), which forms concave and convexly curved exteriors (Figure 5) that are flexible (87). These properties in combination with the variation in amino acids throughout the tandem arrays of TPR motifs are proposed to allow for a diverse binding of ligands, which often occurs at the concave surface but can infrequently occur along the convex surface, as well (83).

Proteins that consist of TPRs can bind a diverse group of ligands in alternate binding locations (83). Ligand binding is believed to be specific and can occur within a populated cellular environment. To accomplish this, not only do individual residues along a particular surface contributing to binding (*eg.* charge and hydrophobicity differences can attract separate ligands), but multiple distinctive TPR folds are used as interaction platforms that can present multiple interfaces for specific binding with ligands (83). Binding of TPR proteins to their ligands has been suggested to be achieved by multiple complex factors, including hydrophobic pockets, charge, amino acid type, hydrogen bonding, electrostatics and coordination of surfaces (83). Although characteristics of TPR binding have been studied, it is still difficult to predict how specific TPR motifs will bind a specific ligand. This problem is largely due to the fact that there is a dearth of TPR structures containing bound ligands in the protein structure databases.

Part of the reason for this may be due to the fact that X-ray crystallography is a protein structure determination method that relies on ordered proteins, which may be problematic with the amount of conformational flexibility inherent in TPR binding to ligands (83). Although flexibility within a TPR protein is not often evident through examination of the crystal structure, one example was recently found that illustrated the extreme flexibility of a TPR that can exist in multiple conformations, due to hinge regions (87). Hinge regions have been found to exist between individual TPR motifs that provide the ability to change conformation (87). For example, in MamA, a hooked shaped magnetosome-associated TPR containing protein, the linker region provides enough flexibility for binding a second TPR domain, which induces a change in conformation of the peptide to a helical state. This novel function provides evidence that TPR proteins can exhibit considerable elasticity that may contribute to an abundance of different functions across diverse species.

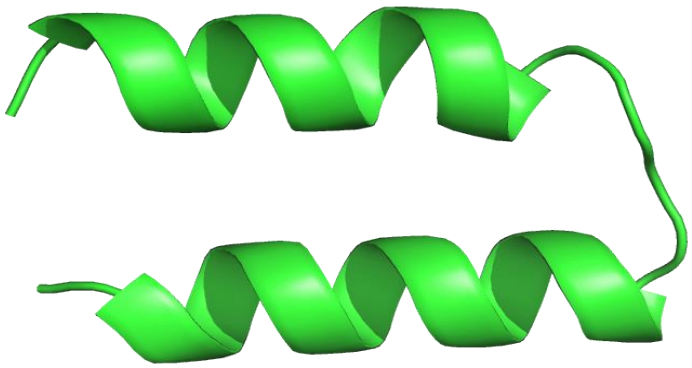


Figure 4. TPR Motif Representation. TPR proteins characteristically contain a basic helix-turn-helix fold of a duplicated, degenerate 34 amino acid sequence. The N-terminus is at the top and the C-terminus are at the bottom.

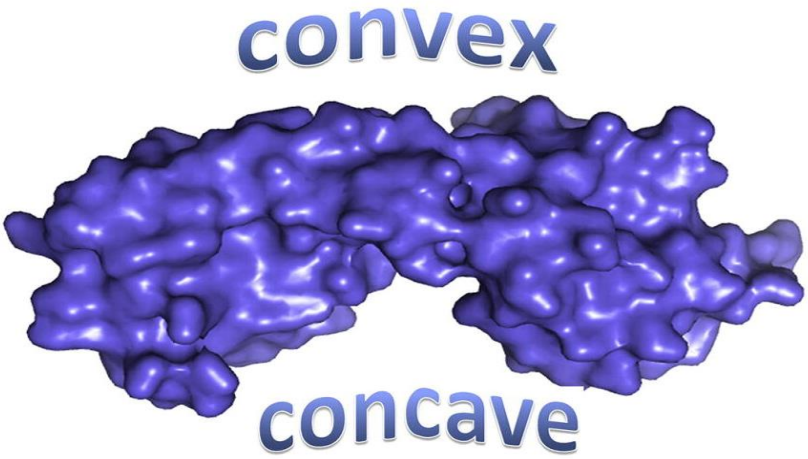


Figure 5. TPR Motif Surface Structure. Convex and concave surfaces of a TPR-containing protein (modified from (88)).

1.7 BcsC

The cellulose export protein BcsC, the protein of interest for this thesis research, has a predicted involvement in the transport of cellulose across the outer membrane (74). Genetic mutants involving BcsC in *Acetobacter xylinus* have indicated that BcsC is required for cellulose synthesis *in vivo*, but only BcsB is required for cellulose synthesis *in vitro* (89). One reason for this may be that cellulose is not being exported. In the *Salmonella typhimurium* MAE52 strain, cellulose was produced, yet mutating BcsC resulted in a severe reduction in biofilm formation and no expression of cellulose (66). BcsC is comprised of over 1,100 amino acids and is predicted to contain a C-terminal 18-stranded β-barrel region in the outer membrane proceeded by an extensive N-terminal TPR in the periplasm. The TPR region of BcsC is predicted to contain between 18 and 21 TPRs, which contrasts with AlgK (10 TPRs) and PelC (19 predicted TPRs) from other EPS biosynthetic systems. The reason for this discrepancy in TPR region length between the analogous proteins remains unknown, yet other key differences exist between their respective pathways. For example, the alginate pathway is believed to have several more proteins in the periplasm than the cellulose pathway. A number of these proteins (*i.e.*, AlgX,

AlgG, and AlgL) have been proposed to assist in periplasmic passage of alginate in conjunction with AlgK, which may act as a scaffold for the assembly of these proteins (74). However, in the Bcs system there are far fewer accessory proteins to assist BcsB-BcsA and BcsC with the passage of cellulose through the periplasm (Figure 2) (77). This may partly account for the increase in the amount of TPRs in BcsC compared to AlgK, yet the specific mechanisms of the structure and function of BcsC remain largely uncharacterized to verify this theory.

Recently, a study of BcsCs TPR region cloned from *Enterobacter* CJF-002, yielded the structure of $BcsC^{TPR1-6}$ (Asp24-Arg272) solved to a resolution of 3.27Å and presented small angle X-ray scattering (SAXS) data of $BcsC^{TPR1-17}$ Asp24-Leu664 (90). The crystal structure of $BcsC^{TPR1-6}$ was helical with 14 α-helices, 12 of which formed TPRs α1–α2, α3–α4, α6–α7, α8–α9, α10–α11, and α12–α13, while the other 2, α5 and α14, were not believed to belong to TPR motifs (90). The crystal contained five monomers in the asymmetric unit and three different conformations were observed. Following superimposition of the five monomers, each of the three conformations showed differences at the turn region (between α5–α6) and in the C-terminal half of each TPR (α6–α11), which extend in different directions (90). This apparent flexibility in structure was suggested to have a hinging effect on the C-terminal super-helix and the effect of this non-TPR region was proposed to assist in changing the direction of the super-helix (90).

When discussing the SAXS data presented by Nojima and colleagues, it is important to note that Asp24-Leu664 was designated as $BcsC^{TPR1-17}$, yet their bioinformatics predictions suggested that the full length TPR region of BcsC contained 19 TPR motifs (90). This differs from the preliminary Weadge lab analysis of BcsC which indicated 18 TPR motifs, it is unclear whether this was due to a difference in the genome of *Enterobacter* CJF-002 and *E. coli* K12 or differences in the analyses. When parsing the SAXS and crystallography data, Nojima and

colleages proposed that BcsC contains 6 super-helices in total that are connected by five hinge/non-TPR regions (90). The non-TPR regions were suggested to be involved in directional changes of the super-helices and might allow the BcsC TPR regions to form a unique structure aiding in the transport of emerging cellulose chains (90). Due to the discrepancies in TPR regions, and the resolution obtained from their model, many questions remain with respect to the structure of BcsCs TPR region in addition to the possible interactions this domain may have with neighboring periplasmic regions of BcsZ and BcsA-B.

2. Research Need

Bacterial cellulose synthesis and export processes is significant to many bacterial species and is integral to biofilm formation and propagation in these organisms. Although bacterial cellulose research has expanded into many different areas, such as food and food packaging, cosmetics, drug design, and drug delivery systems (61), the primary implications for bacterial cellulose research, with respect to this thesis, is found in the medical field due to its role in pathogenesis. For example, bacterial biofilms are utilized by many bacteria that have been implicated in chronic and persistent infections (15, 16). Furthermore, biofilms can provide a mechanism to survive in harsh environments in order to inflict pathogenicity on the host and lead to a variety of ailments, including cancer (4, 16, 91). Despite this, insufficient information regarding their exopolysaccharide biosynthetic structures and mechanisms exists. Worldwide foodborne diseases are estimated to infect 600 million people annually, resulting in 420,000 fatalities (1). The United States alone averages 48 million cases each year, causing a predicted economic burden of 77.7 billion in health-related costs (2). Recently in North America, outbreaks of *E. coli* have been linked to beef, flour, and produce (10, 92, 93), with some strains that have proven resistant to many varieties of antibiotics (93). A similar situation exists with *Salmonella* outbreaks that have been reported in chicken, turkey, and even cereal (94–96). As it seems there is no end in sight, additional research is crucial to alleviating the constant pressure caused by these bacteria both medically and economically, as the majority of bacteria exist within biofilm communities (97, 98). In addition, the study of bacterial cellulose proteins is of paramount importance to our food and water security and can consequently provide improved knowledge of biofilm production as many bacterial pathogens, such as *Salmonella* and *E. coli* employ biofilms as a protection strategy (99–101).

Recently, many molecular insights into the polymerization and transport of bacterial cellulose have been made (76–78, 90, 102). However, little is known of how the final steps in export of the polymer occur. BcsC is a key cellulose export protein in this process, believed to be essential for proper biofilm formation (66, 90) and, thus, presents a significant target for preventing and controlling biofilm production, bacterial expansion, and disease progression. BcsC is a member of a large class of exopolysaccharide export proteins (including AlgK), which are inherent components of biofilm biosynthetic apparati from species across the bacterial kingdom. Thus, a greater understanding of the role of BcsC is likely to have widespread implications in the control and disease prevention of many pathogenic bacteria, but also possibly in promoting the colonization of potentially beneficial bacteria (*eg*., probiotics). Current bioinformatical knowledge and research on BcsC in the Weadge lab has indicated that BcsC interacts with cellulose and predicts that BcsC may contain partially hydrophobic character that could theoretically be involved in ligand binding (previous work by Emily Wilson and Alex Anderson). *From this preliminary data,* ***we hypothesize that structural investigation of the N-terminus of BcsC will confirm that it contains a series of TPR folds and that this region is important in ligand binding.*** These results will ultimately provide further insight into the role of BcsC in cellulose export. For example, a structural model will aid in the identification of possible substrate binding regions, hydrophobic interactions, and/or protein-protein interaction surfaces that can be further explored in additional functional and phenotypic experiments.

To address our hypothesis, we will perform the following objectives:

Objective #1: Protein expression and purification of the series of protein constructs (Table 1 results section) of the BcsC TPR domain.

Objective #2: Perform extensive crystal screening of the TPR protein constructs for conditions that stabilize the purified protein constructs (dynamic light scattering) and are amenable to protein crystal formation. Promising conditions will then be optimized to refine the crystallization process so that the resulting crystals are suitable for X-ray diffraction data collection.

Objective #3: Build structural models of the protein constructs of BcsC through a combination of X-ray crystallography and small-angle scattering experiments.

3. Materials and Methods

3.1 Reagents, Chemicals, and Media

The chemicals purchased from Fisher Scientific were EDTA-free protease inhibitor, sodium chloride, tris(hydroxymethyl)aminomethane (Tris). In addition, Coomassie stain and destain were prepared using chemicals also obtained from Fisher Scientific. The products purchased from BioBasic were BcsC synthetic gene constructs, Tryptone powder, isopropyl β-D-1-thiogalactopyranoside (IPTG), and yeast extract. D-glucose, imidazole, and RNaseA were obtained from BioShop. DNaseI was purchased from Fermentas and Q-Sepharose Hi-Trap columns were purchased from GE Biosciences. Through a purchase from BioRad, we obtained 5mL IMAC columns and Nickel-nitrilotriacetic acid chelate (Ni-NTA) resin was purchased from Quiagen. Vivaspin centrifugal ultrafiltration devices were obtained from GE healthcare. The crystal screens Top 96, MCSG-1, 2, 3, and 4 were purchased from Microlytic and the screens purchased from Molecular Dimensions were Morpheus BN201-1-47, JCSG-Plus BN216-1-40, and PACT Premier BN163-1-36. The 96-well INTELLI-PLATEs were purchased from Art Robbins Instruments and Crystal Clear sealing tape and IZIT dye was obtained from Hampton Research. In addition, 24 well plates were purchased from Crystalgen and the SelenoMet Nutrient Mix and Medium Base was purchased from Molecular Dimensions. Any reagents or chemicals not listed above were purchased from Sigma-Aldrich. The media Luria-Bertani (LB) is commonly used to grow stock cultures after transformations and is also used for general culturing of the bacterial strains. LB was made with tryptone (10 g/L), yeast extract (5 g/L), and NaCl (10 g/L). Solid LB media when used contained 1% (w/v) agar. A nutrient rich medium, termed Super Broth (SB), was used for protein production with bacterial cultures and contained

tryptone (32 g/L), yeast (20 g/L), and NaCl (10 g/L). When selecting for target plasmid, kanamycin sulfate was added to the media at a final concentration of 50 µg/mL.

3.2 Bioinformatics Analysis

Bioinformatics analyses were conducted on the full amino acid sequence of BcsC and all sub-constructs prior to the onset and throughout this project to gain a familiarity with the hypothesized results and track any changes that may have occurred as databases were modified. Briefly, an initial examination of protein characteristics was conducted using the ProtParam software (103) that provide information regarding the theoretical isoelectric point (pI), molecular weight, molecular extinction coefficient and the number of specific amino acids present in each of the protein constructs. This information was crucial for checking protein concentrations, optimizing the pH for buffers and other experiments. Additionally Phyre^2, a predictive and analytical tool for both structural and functional aspects of proteins (104), was used to generate sequence homology models with other proteins in the protein databank. The generated hypothetical protein databank (PDB) files from Phyre^2 were used for figures following graphical rendering in PyMOL (105). Multiple sequence alignments of BcsC were generated using protein sequences from various organisms via the program BLASTP with the built in iteration PSI-BLAST (106, 107) and with Clustal omega (108). Meta disorder prediction programs were also utilized to identify potential disorder protein regions (109).

3.3 Protein Constructs and Cloning

The protein BcsCs TPR domain sequence had been identified in previous bioinformatic searches (UniProt P37650) from the sequenced geonome of *E. coli* K12. Prior to the start of this project Phyre^{2}s secondary structure predictor was crossed checked with TPRpred and disorder prediction programs to formulate the sequences to create the template for the constructs. Six

protein construct derivatives that together cover overlapping portions of the full TPR region of BcsC (Figure 6) were codon-optimized to increase the chances of crystallization and to investigate ligand/substrate binding in different regions of the protein. For example, certain TPR sections may have a stronger affinity for the substrate cellulose and/or certain proteins may stack more uniformly for crystallization purposes. These constructs were designed by Dr. Joel Weadge and given specific naming designations based on the represented TPR regions, as $BcsC^{TPR1-8}$, $BcsC^{TPR1-11}$, $BcsC^{TPR1-15}$, $BcsC^{TPR4-21}$, $BcsC^{TPR9-21}$, $BcsC^{TPR12-21}$, with 1 representing the N-terminal TPR motif and 21 representing the C-terminal motif. Each construct was subcloned by BioBasic into its own pET28a expression vector, which was tailored for protein expression using the isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible T7 promoter that is under control of the Lac operon operator. IPTG is analogous to lactose in function and when present it removes the lac repressor (LacI) from LacO. Transcription of a gene of interest downstream from the T7 promoter ensues once the cell-encoded T7 RNA polymerase binds. Each expression vector is also designed so that transcription/translation results in the inclusion of a His_6-tag to the respective N or C terminal end (see Table 1 Results section for specific constructs) of the protein to facilitate downstream purification and detection steps.

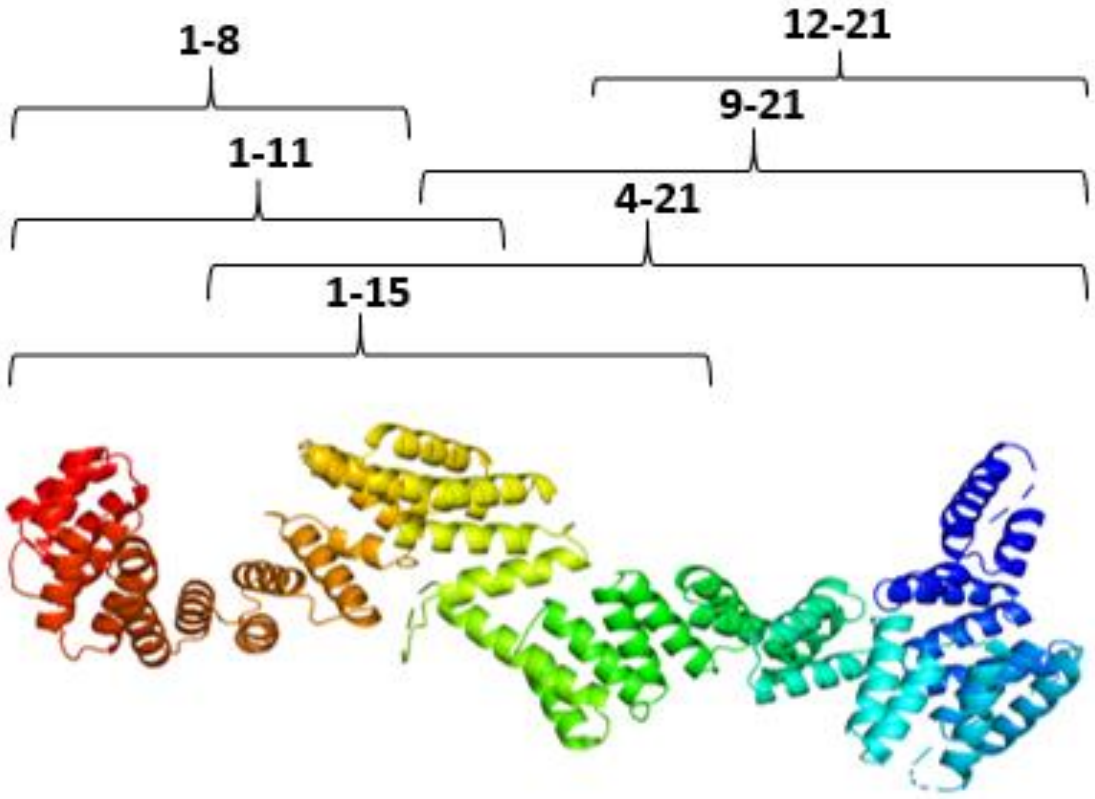


Figure 6. Visual Display of BcsCs TPR Constructs. All six protein constructs are illustrated with estimated length overlaid on a sample TPR image. Note that this is only a representative image and protein constructs and TPR image is not fit to scale.

3.4 General Expression and Purification Strategy

For the subsequent sections see the attached flow chart to aid in visualization of methods (Figure 6). Plasmids containing specific protein constructs with antibiotic resistance were individually transformed into $CaCl_2$ *E. coli* BL-21 competent cells (Novagen) using the standard heat shock method (110). Briefly, 5 µL of pET28a plasmid was dispensed into a standard 1.5 mL microfuge tube which contained 100 µL of BL-21 competent cells (Novagen) and was incubated at approximately 4°C for 30 min. The sample was then heat-shocked at 42°C for approximately 1.5 min, placed back at 4°C for 5 min before aseptically adding 500 µL pre-warmed (37°C) LB

broth and incubating at 37°C for 1 h with shaking at 200 rpm. Transformed cells were plated on LB agar (supplemented with 50 µg/mL kanamycin) and incubated for 14-16 h at 37°C.

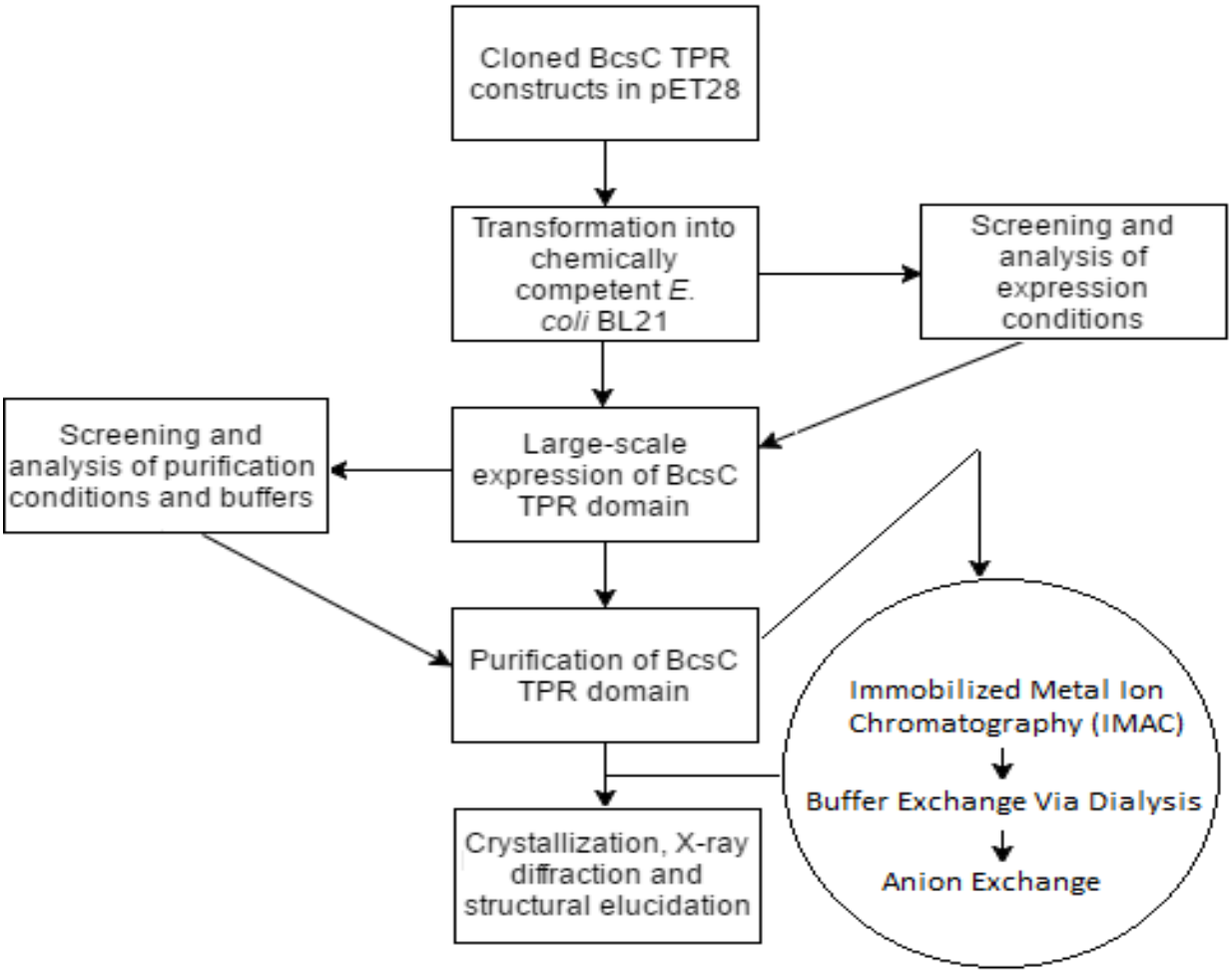


Figure 7. **Flow Chart of Experimental Procedures**. This flow chart contains a step-by-step outline of the methods section. The flow chart illustrates the streamlined expression and purification steps in the middle and emphasizes the extensive optimization of elements to either side.

3.4.1 Protein Expression

Protein expression was conducted in a large-scale fashion with the objective of overproducing recombinant protein from each respective BcsC construct transformed into *E. coli* BL21 (pET28) cells. First, initial stock cultures of transformed cells were created by inoculating 5 mL LB (supplemented with 50 µg/mL kanamycin) with transformed cells followed by

incubation at 37°C for 15-18 h with shaking (200 rpm). Stock cultures were then used at a ratio of 1/50 to inoculate 1 L cultures of SB (supplemented with 50 µg/mL kanamycin). The SB cultures were then incubated at 37°C with shaking (200 rpm) until the optical density at 600nm (OD_{600}) measured between 0.6-0.8 and then IPTG was added to a final concentration of 0.5 mM. Following addition of IPTG, induction of protein expression was allowed to continue at 22°C for 8-16 h (with shaking at 200 rpm) until the cells were harvested by centrifugation (5000 x *g* for 15 min at 4°C). The supernatant was discarded, and the collected cell pellets were stored at -20°C until needed.

3.4.2 Protein Purification

Several purification steps must be completed to obtain pure protein, as required for downstream applications (111). These are outlined below in detail successively in the order which they would be performed during a typical purification.

3.4.2.1 Cell Lysis: Frozen pellets containing the equivalent of 2 L of culture were re-suspended in 40 mL of lysis buffer (50 mM Tris-HCl pH 8, 300 mM NaCl) to which RNaseA at a concentration of 0.0125 mg/mL, DNaseI at a concentration of 0.025mg/mL, and one Pierce protease inhibitor tablet was added. The suspension was thawed on a rotating agitator at 4°C until the solution was homogenous. Cell lysis was achieved using one pass through a standard cell disruptor (Constant Systems TS Series 0.75kW machine Pressure Biosciences) operating at 17 kpsi sample pressure. The lysate was centrifuged at 20,000 x *g* for 45 min at 4°C to separate the soluble protein from the cellular contaminants (*i.e.*, whole cells and inclusion bodies), the supernatant was collected, and the pellet was discarded.

3.4.2.2 Immobilized metal ion affinity chromatography (IMAC): Approximately 2 mL of settled nickel nitrilotriacetic acid (Ni-NTA) resin beads was flushed with approximately 50

mL of lysis buffer to remove the ethanol storage buffer. The beads were then suspended in the cleared lysate 4°C for 1-2 h on a rotating agitator to facilitate binding of recombinant protein to the Ni-NTA resin beads. The resulting solution was then applied to a gravity-flow 50 mL column connected to diastolic pump (set at a constant pressure of 1.5 mL/min throughout the procedure) and the flow through from the column was collected. The resin remaining in the column, containing the His_6-tagged proteins of interest, was washed with 50 mL of lysis buffer, followed by 50 mL of wash buffer I (50 mM Tris-HCl pH 8, 300 mM NaCl, 20 mM imidazole), and 25 mL of wash buffer II (50 mM Tris-HCl pH 8, 300mM NaCl, 40 mM imidazole). Finally, the column was washed with 25 mL of elution buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 250 mM imidazole), which was left to incubate for 5-10 min before collection and storage at 4°C. The purpose of this method was to release contaminants in the first 3 washes, while losing only minor amounts of the target protein, leaving the final elution consisting of primarily the desired recombinant BcsC protein. A standard sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was conducted to determine protein purity (as outlined below).

3.4.2.3 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE):

Samples were prepared by combining 40 μL from each of the fractions collected (flow through, wash and elution from chromatography columns) with 20 μL of 5 times concentrated SDS sample buffer (1 mL 2 M Tris-HCl (pH 6.8), 5 mL glycerol, 1.0 g SDS, and mL 0.2% Bromophenol blue brought to 10 mL water MQH_2O) with 1 mM dithiotreitol (DTT) and heated for 5 to 10 min at 90°C. The Mini-PROTEAN tetra cell apparatus (Biorad) housed the gels and was filled to the appropriate volume with running buffer (28.8 g glycine, 6.04 g Tris, 2 g SDS, and 1.8 L dH_20). The first lane was aliquoted with 7 μL of precision plus dual colour protein standard (BioRad) as a molecular weight reference and each subsequent well had 15 μL sample

dispensed into it. The only variation of this process was conducted for expression gels that involved taking 0.5 mL of culture, pelleting it at 10,000 x *g* and resuspending in 50 µL of SDS. After heating as described above, 20 µL (for T=0; induction), 15 µL (for T=1; ~4 hr), and 10 µL (for T=2; ~16 hr) were added to each subsequent well to balance protein levels in favour of a more accurate measurement. After the apparatus was closed, the gels were run at 200 V for 45 min, followed by staining with Coomassie R250 stain solution (2 g Coomassie Brillant Blue R250, 500 mL dH_20, 400 mL methanol, 100 mL acetic acid (glacial) or transferred to nitrocellulose paper for Western blotting analysis. Stained gels were heated in a microwave for 30 s to speed up the staining process and left to sit on a rocking shaker for 20 to 40 min. Afterwards, Commassie R250 stain solution was decanted, de-stain solution (700 mL dH_20, 200 mL methanol, 100 mL acetic acid (glacial) was added, and the process of heating and rocking was repeated every 20 min until the desired contrast was achieved.

3.4.2.4 Western Blots: Western blot transfers were completed using a Trans-Blot apparatus (Biorad) filled with transfer buffer (12 mM Tris-HCl, pH 7.5, 96 mM glycine, 20% (v/v) methanol) and run at 4°C for 2 h at 100 V. Once transfer onto nitrocellulose was completed, the blots were then blocked with blocking buffer (5% (w/v) skim milk powder in TBS buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl)) for approximately 1 h. The blots were then washed twice in TTBS (20 mM Tris-HCl, pH 7.5, 2 mM NaCl, 0.05% (v/v) Tween) for 7 min each. Primary antibody (mouse anti-His) was added to a dilution of 1000-fold in 15 mL blocking buffer and incubated for 45 min with the blots. After primary incubation, the blots were washed in TTBS for 7 min three separate times followed by a 45 min incubation with a 5000-fold dilution of secondary antibody (alkaline phosphatase conjugated rabbit anti-mouse) in 15 mL of blocking buffer. Lastly, the blots were washed once more in TTBS. The treated

nitrocellulose paper was developed by lightly coating the surface in 5-bromo-5-chloro-3-indolyphosphate (BCIP) substrate solution (BioShop) and incubated in the dark for 5 to 10 min. When the alkaline phosphatase cleaves BCIP on the conjugated secondary antibody, nitroblue tetrazolium (NTB) is formed leaving a purple precipitate localized on the His_6-tagged target protein.

3.4.2.5 Dialysis: Dialysis was used as the preferred method of buffer exchange for all experiments. Dialysis was used to gently remove salt, imidazole, and possible contaminants before further purification steps. Collected protein was transferred into dialysis tubing (2 cm of dialysis tubing/1 mL of sample) that had been cut to fit the total volume of sample (the average was approximately 25 mL) and briefly soaked in distilled water (to increase pliability). After the sample was secured by knots in the tubing and clips on the ends, the sample was submerged in 2 L of dialysis buffer and left at 4°C to gently mix with a stir bar for 1 h. This process required replacing the initial dialysis buffer with a new 2 L volume of the same buffer for another 2 to 3 h. The dialysis buffer consisted of 50 mM Tris-HCl pH 8 for all constructs, excepting $BcsC^{TPR\ 1\text{-}11}$ that required 50 mM Bis-Tris pH 6 for cation exchange experiments.

3.4.2.6 Ion Exchange: The principle behind the second chromatographic step was to separate the protein of interest from contaminant proteins based on charge. Two types of ion exchange chromatography exist: anion and cation exchange. Anion exchange uses a positively charged resin that attracts negatively charged molecules and was suitable for all protein constructs with a pI <7. Cation exchange is the opposite and was more suitable for proteins with a pI >7, such as $BcsC^{TPR1\text{-}11}$. The ion exchange was run using a GE Akta Pure FPLC instrument with a GE Hi-Trap 5 mL Q-Sepharose anion exchange column. The elution strategy was similar to the IMAC procedure, except salt was used instead of imidazole to elute the protein. Protein

samples were passed over the column twice at a rate of 5 mL/min to promote binding to the resin. The column was then washed in dialysis buffer (50 mM Tris-HCl pH 8 for anion exchange and 50 mM Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane (Bis-Tris) pH 6 for cation exchange) at a flow rate of 5mL/min and eluted with a gradient of dialysis buffer and anion exchange elution buffer (50 mM Tris-HCl pH 8, 1M NaCl) or cation exchange elution buffer (0.5 mM Bis-Tris pH6, 1 M NaCl) for $BcsC^{TPR1-11}$. The gradient consisted of a gradual shift from 0% to 100% of cation exchange elution buffer at a rate of 2% per min. Protein constructs eluted differently based on their predicted pI, and the progress of elution and collection of fractions was monitored at $Ab260_{nm}$ while the fractions were collected in the GE Akta fraction collector. After ion exchange was completed, SDS PAGE was conducted using the fractions containing protein to verify the presence and purity of the desired protein.

3.5 Dynamic Light Scattering

Polydispersity was analysed using Dynamic Light Scattering (DLS). The polydispersity of the protein can indicate how stable the protein is within a particular buffer (112). A low polydispersity level indicated that the protein was not aggregated and that the protein was pure. A high polydispersity level may have been an indication of contaminant proteins or aggregation among even the purest protein, due to unsuitable buffer conditions and/or precipitation. Fractions from the anion/cation exchange containing the protein of interest were collected and analysed through DLS. The DLS samples were first filtered in microcentrifuge tubes and then dispensed into 35 µL wells in duplicate using protein concentrations that ranged from 0.5 mg/mL to 5 mg/mL and performed in 50 mM Tris-HCl 150 mM NaCl). Data was collected at room temperature (approximately 25°C) with 15 acquisitions at 5 s intervals via a DynaPro plate

reader (Wyatt Technology, Santa Barbara, CA), and then analyzed using Dynamics software (version 7.1).

3.6 Protein Concentration

Prior to being placed into crystal plates the protein was first concentrated to maximize the chances of crystal formation (111). All protein constructs were collected after ion exchange and concentrated using an ultrafiltration apparatus (Centricon) with a 30 kDa molecular weight cut-off filter; excepting $BcsC^{TPR1-8}$ that used a 10 kDa filter cutoff. Protein concentration was quantified at A_{280} with the respective extinction co-efficient for each protein construct.

3.7 Crystal Screening Trials and Optimization

Crystal screening trials with protein constructs $BcsC^{TPR1-15}$ and $BcsC^{TPR12-21}$ were completed using several different screens (Morpheus, PACT Premier, JCSG-plus, Top 96, MCSG-1, 2, 3, and 4), at a range of different protein concentrations (10-40 mg/mL). The Gryphon robot (Art Robbins Instruments) was pre-programmed for different protein to buffer ratios (2:1, 1:1, and 1:2) of between 0.2 µL and 0.6 µL of protein in each drop and with 45 µL of buffer per well in sitting drop 96-well plates. All sitting drop 96-well plates were incubated at 18°C and 24-well expansion plates were incubated at either 4°C, 18°C, or 20°C and periodically checked by microscopy (Olympus SZX16 Stereomicroscope). Initial crystal screening trials yielded crystal hits that were screened for false positives (salt crystals) using ultraviolet light or IZIT protein dye. When assessing with ultraviolet light, protein crystal fluorescence due to the excitation of aromatic residues (*ie.* tryptophan) within the protein, whereas salt crystals do not. Similarly, when staining with IZIT dye protein crystals absorb the purple dye and turn a dark shade of purple or blue, whereas, salt crystals would not.

Once a promising crystal hit was discovered, crystallographic expansion trials were conducted by recreating the initial crystal growth condition on a larger scale using pre-greased hanging drop 24 well crystal plates. Protein and buffer were carefully aliquoted onto 22 mm siliconized glass cover slides before each slide was placed on top of each well which contained 500 μL of buffer to generate a hanging-drop crystallization chamber. A range of different ratios were used as the volume (in μL) of protein to buffer was varied from 4:1, 4:2, 3:2, 2:2, 2:1, 2:2, 1:2, 2:3, 2:4 and 1:4 depending on the trial. The preparation of buffers for the crystal conditions involved creating a stock solution that was balanced to the proper pH and then diluted to the proper concentration, followed by the addition of salts, and/or additives such as PEG or glycerol. For example, optimization of the MCSG-3 G12 condition (0.1 M sodium acetate pH 4.6, 2 M sodium formate) involved varying the buffer (sodium acetate) concentration in one row while holding the precipitant (sodium formate) constant, then varying the precipitant concentration in another row while holding the buffer concentration constant. Following this, variations of both precipitant and buffer concentration not yet attempted were prepared in a grid style screening. The final pH of the condition was always checked against the pH of the premade screen from which it was derived from to ensure consistency.

A variety of additives (glycerol, DMSO, and ethylene glycol) were used in conditions that already produced crystals, with the hopes of refining the crystals, at a variety of different concentrations (1.25, 2.5, 5, and 10% (v/v)). In addition, some trials were completed with the substrates D-glucose and cellobiose pre-incubated with the protein in co-crystallization efforts. Furthermore, to increase the chances of crystallization, a crystal streak seeding technique was used to introduce nucleation sites (113). This technique was used in combination with additives, as well as substrates, with the intention of optimizing pre-existing conditions. Crystal seeding

can be effective when placing crystal seeds into a supersaturated zone, termed the metastable zone, because growth occurs readily in this zone, but nucleation points do not form and hence the introduction of nucleation points can effectively allow crystals to grow (Figure 8) (113). Crystallization without seeding involves creating a condition that begins in the supersaturated labile zone (nucleation points can form) followed by the supersaturated zone to promote growth. However, this process can yield crystals that are difficult to harvest due to precipitation surrounding the crystals (113). The common concern with crystal seeding is that an abundance of nuclei will be placed into the supersaturated solution and yield masses of crystals unsuitable for diffraction analysis (*ie.* typically too small) (113). When this was observed for seeding in our crystal growth experiments, a dilution series was created from the seed stock and a more suitable fold dilution was used (10^2 or 10^3).

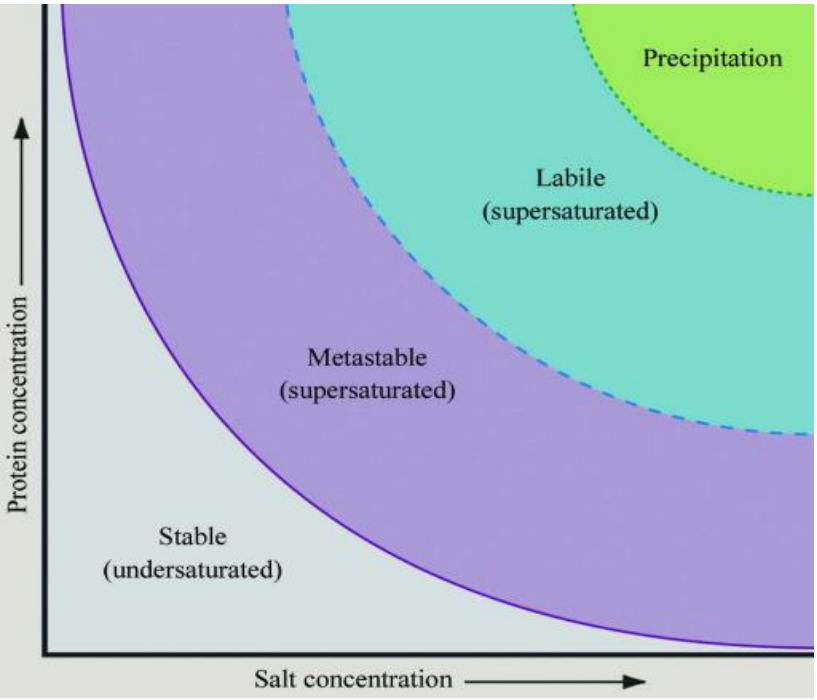


Figure 8. Phase Diagram for Crystallization. The y-axis represents the protein concentration and the x-axis represents salt or precipitant concentration. The stable, undersaturated zone usually represents a clear well with no crystallization. The metastable supersaturated zone can develop nuclei into crystals, yet nucleation points do not form here. The labile supersaturated zone can both form nucleation points and support crystal development. The supersaturated precipitation zone is often a region that contains precipitation and can support crystal growth, though may not yield easily harvestable crystals (adapted from 94).

To facilitate uncovering optimal screening conditions in a limited time frame for this thesis, proteins $BcsC^{TPR\ 1\text{-}11}$, $BcsC^{TPR\ 9\text{-}21}$, and $BcsC^{TPR\ 12\text{-}21}$ were sent to the Hauptman-Woodward Medical Research Institute (HWI) for high-throughput crystallization screening. This facility performed 1,536-well microassay plate screenings for the BcsC protein constructs using a microbatch-under-oil technique (114). HWI used automated liquid handling to facilitate set-up of the crystal screens and each well was imaged before adding the protein solution, after addition of the protein solution, followed by imaging at the one day, one week, two week, three week, four and six week timepoints. Only the final well images are accompanied by Second Order

Nonlinear Optical Imaging of Chiral Crystals (SONICC) images from a Formulatrix Rock Imager 1000 designed to determine if an object (as small as <1 µM) was crystalline (115) and images from two-photon excited ultraviolet fluorescence (TPE-UVF) to further interrogate if it is protein (116). HWI screening with BcsC constructs was performed twice. For the first experiment, protein samples were prepared the night before screening, shipped express overnight, and stored at 4°C before screening, while the second trial involved freezing the samples at -80°C and shipping overnight express on dry ice in an attempt to preserve protein quality.

3.8 X-ray Diffraction

Crystals that were selected for X-ray diffraction were analyzed either at the University of Waterloo or the Canadian Light Source. Prior to analysis, crystals were looped using an appropriate size loop (Mitegen) and soaked in a cryo-protectant solution consisting of the mother liquor (condition in which the crystal was formed) supplemented with either 33% (v/v) glycerol or ethylene glycol (for crystals from Top 96 A1: 0.2 M MgCl, 0.1 M Tris-HCl pH 8.5, 30% (w/v) PEG 4000), or 6 M sodium formate (for crystals from condition MCSG-3 G12: 0.1 M sodium acetate pH 4.6, 2 M sodium formate) for 20-60 s. The crystals were vitrified in liquid nitrogen and stored in liquid nitrogen until diffraction was performed. Full crystal data sets were collected using synchrotron radiation on the 08B1-1 beamline at the Canadian Macromolecular Crystallography Facility (Canadian Light Source, Saskatoon) using a CCEL MD2 microdiffractometer and MarMosaic mx300 CCD X-ray detector. Typical datasets consisted of 360 images at 1° oscillations and an exposure time of 0.2 s per image at the CLS. The data were integrated, reduced and scaled using AutoProcess on *MxLive* (117).

3.9 Molecular Replacement and Model Building

Molecular replacement was attempted to determine the phases of the collected data. PDB files from previously solved structures of $BcsC^{TPR1-6}$ (PDB ID: 5xw7) and AlgK (PDB ID: 3e4b), as well as structures with at least 25% amino acid sequence identity to $BcsC^{TPR1-15}$ identified with Phrye[2] and BLASTP searches were used as possible replacement models. These replacement models were prepared by using the Sculptor (118) of PDB tools program in the Phenix suite (119) to edit the PDB files to not contain heteroatoms and consist of a single polypeptide chain. In later steps, further trimming of the PDB files to alanine traces of the model was also accomplished with Sculptor. Molecular replacement with each of these templates was then attempted with the MRage and/or Phaser modules (120) of the Phenix suite. As an alternative, the automated molecular replacement was also attempted online through the CCP4 online interface (121) with the BALBES (122), MrBump (123), and MoRDa (124) programs. Following molecular replacement by these methods, model building was attempted with Autobuild in the Phenix suite through iterative variations consisting of different rounds of refinement, building helices and strands and morphing the input model into density features.

3.10 Selenomethionine Expression

As an alternative route to solving the structure with molecular replacement, selenomethionine (SeMet) derivatives of the protein were generated to employ anamolous dispersion techniques for structure determination. SeMet media was prepared by mixing 21.6 g SeMet Medium Base with 1 L of ddH_2O and autoclaving for sterility. Prior to use of the media, kanamycin was added to a final concentration of 50 μg/mL along with Nutrient solution (5.1 g SeMet Nutrient mix) premixed in 50 mL of MQH_2O and sterilized by passing through a 22 μm filter (VWR) and a final 40 μL concentration of analytical grade L-(+)-SeMet. Prior to

inoculation of the SeMet media, a 50 mL stock culture grown in LB media (as opposed to 20 mL used for native expression) was pelleted by centrifugation (4000 x *g*, 20 min, 4°C) and carefully rinsed once with MQH_2O to eliminate traces of methionine. The culture pellet was resuspended in a minimal volume of sterile SeMet media then aseptically added back to the larger 1 L of SeMet growth medium. SeMet expression of protein construct (only $BcsC^{TPR\ 1\text{-}15}$ was expressed thus far) was then conducted under the same conditions as native expression (22°C, 200 rpm). When the culture reached an OD_{600} of 0.6, IPTG was added to a final concentration of 1 mM to induce expression of the target protein, and the culture was allowed to incubate for an additional 16-18 h, followed by centrifugation at 5000 x *g* and the pellet stored at -20°C until needed. All techniques for protein purification were the same as outlined in previous sections. SeMet prepared versions of $Bcs^{TPR1\text{-}15}$ required anion exchange purification after IMAC and dialysis. The $Bcs^{TPR1\text{-}15}$ SeMet crystal expansion plates were set up in a condition containing 0.1 M sodium acetate pH 4.6, 2 M sodium formate at a concentration of 25 and 28 mg/mL.

3.11 Small angle X-ray scattering (SAXS)

SAXS data was collected at the SIBYLS beamline (12.3.1) at the Advanced Light Source part of the Lawrence Berkeley National Laboratory (125). The X-ray wavelength used was 1.0 Å with a flux of 1013 photons per second and the sample-to-detector distance set to 1.5 m. Scattering images were collected using a Pilatus 2 M detector every 0.3 s, with a total of 33 images per sample. All sampling was performed at 10°C and data was processed as described (126). Briefly, a collection of three separate protein concentrations (1, 5, and 10 mg/mL) was used to correct for concentration-dependent behaviour and two protein-free buffer samples were collected, for every 3 sample concentrations, to reduce error in subtraction (127, 128). Every

collected image was circularly integrated then normalized for beam intensity to generate a 1-dimensional scattering profile (127, 128).

The 1-dimensional scattering profile of each respective protein sample was buffer-subtracted by the two respective buffers to produce two sets of buffer-subtracted sample profiles (127, 128). Scattering profiles were examined for radiation damage by sequentially averaging them together until radiation damage effects were noticeably altering the scattering curve (127, 128). Averaging was performed using the web-based software program from the SIBYLS website called Frameslice (129). The program SCÅTTER (130) was used to compute the radius of gyration (Rg), Rc, P, Q, and volume parameters, which can be used for corroborating data with current knowledge about the target protein (*ie.* validity checking). The GNOM function (131) was used in the program PRIMUS (132) to compute the pair distribution (P(r) function) and the maximum distance of the molecule (D_{max}) was estimated using the P(r) function.

Imaging was conducted with the program GASBOR (133) through the online data analysis server ATSAS 2.8.4 (https://www.embl-hamburg.de/biosaxs/atsas-online/) (134), which produced models consisting of dummy residues. GASBOR was also used in conjunction with DAMAVER (135) through the SIBYLS beamline website (http://sibyls.als.lbl.gov/) for SAXS applications to analyse the data. The resultant PDB file produced by DAMAVER was opened in PyMOL and fitted with an alanine model of the PDB file (5xw7) from the solved structure of $BcsC^{TPR\ 1\text{-}6}$ using SASpy (136).

4. Results

4.1 Bioinformatic Analyses:

A thorough bioinformatics investigation of the BcsC constructs was conducted using several different programs and was utilized throughout all stages of research. ProtParam (103), a quick and simple tool used, was important for the planning of wet lab experimentation as it provided the molecular weights, extinction coefficients, the theoretical pIs, and other relevant information (Table 1) for calculating protein concentrations, following molecular mass on purification gels and aiding in decisions for purification steps (*eg.* cation exchange for $BcsC^{TPR\ 1\text{-}11}$. Of the constructs generated and analyzed, $BcsC^{TPR\ 4\text{-}21}$ was the largest and encompassed the majority of the TPR region, while $BcsC^{TPR\ 1\text{-}8}$ was the smallest, with only the N-terminal TPR regions included. This range of constructs provided us with a panel of TPR regions and attributes that led to improving our chances of producing soluble purified protein. For example, from the instability index analysis (ProtParam), $BcsC^{TPR\ 1\text{-}15}$ is predicted to be the most unstable by nearly 2 units and rated both $BcsC^{TPR\ 12\text{-}21}$and $BcsC^{TPR\ 4\text{-}21}$ the most stable. However, this may not be significant, as instability index predicts that only proteins over a score of 40 are considered unstable and all BcsC constructs fall within a reasonable range of this limit (103).

Clustal omega, a bioinformatics tool, was used to identify amino acid conservation between BcsC constructs and other similar proteins through a manual input interface. Due to homology, the primary proteins of interest were AlgK and BcsC from Enterobacter CJF-002. The comparison matrix for AlgK is presented from the alignments in Table 2. When examining the sequence identities between AlgK and BcsC constructs, little variation is found. The full-length construct ($BcsC^{TPR\ 1\text{-}21}$) has the exact percentage as $BcsC^{TPR\ 4\text{-}21}$ (24.46%) and $BcsC^{TPR\ 1\text{-}15}$

has the lowest sequence identity (22.51%). The highest sequence identity is BcsC$^{TPR\ 12\text{-}21}$ (25.93%) and the rest of the constructs range between. These percentages may not be high enough to use the AlgK PDB file for molecular replacement, as the recommended search criteria for sequence similarity 25-35% (137). The sequence identity between the full length TPR domain of BcsC from *Enterobacter* CJF-002 and *E. coli* K12 was 72.38, whereas, the individual construct homology (presented in Table 3) ranged between 71 and 73% identity. These results are promising and provide relevance towards SAXS and molecular replacement modelling, for example, a SAXS envelope can be merged with the PDB file from *Enterobacter* CJF-002 in PyMOL that may help signify the validity of the SAXS data. Furthermore, the high sequence identity may provide a strong background to build a molecular replacement model using the phase information from *Enterobacter* CJF-002.

A number of programs were instrumental in aiding our structural work and subsequent hypothesis design. The programs Phyre^2 and PSI-BLAST were used to complement each other in identifying potential homologs for BcsC. PSI-BLAST, a secondary program that builds and refines initial search alignments conducted by BLAST (106, 107), found several homologous sequences and the top 5 results were recorded for later structural modeling (Table 4). Phyre^2 was used to identify homologs (based on amino acid sequence and overall fold) and generate structural models (as Protein data bank (PDB) files) that were compatible with several programs used for tertiary structure modeling/imaging. For example, PyMOL and Coot (138) were used to visualize hypothetical models of the entire TPR domain (BcsC$^{TPR\ 1\text{-}21}$) and 6 of the BcsC protein constructs (Figures 9-15) to visualize individual TPR regions and map possible key conserved residues. The top 5 results of homologous proteins identified by Phyre^2 are presented in Table 5, each individual result represents one protein that has predicted homology to BcsC with a specific

corresponding PDB file (template number) and title of each respective crystal structure. The high alignment coverage is due to amino acid homology because the percent identity is 15% or lower for all models. The structures of these proteins and the BcsC-based models thereof (*eg.* threaded structural models), were used in downstream SAXS and molecular replacement analyses to differing degrees of success (as outlined in subsequent sections).

The program TPRpred (https://toolkit.tuebingen.mpg.de/#/tools/tprpred) (139) was used to predict the number and location of TPRs within the BcsC TPR domain. Prior to this, an analysis of the full length BcsC TPR domain in the secondary structure prediction section of Phyre[2] indicated strong alpha helical character up to approximately residue 790 (See Appendix A1) and from a visual examination of this section counted over 40 alpha helical regions. Accordingly, the BcsC construct designation was extended to allow for the possibility of 21 TPR segments. When the full length BcsC TPR sequence was inputted in TPRpred, 18 TPRs were predicted (Table 6), hence, it is not known the exact number of BcsCs TPR regions, but it was believed to be between 18-21 TPRs.

The final bioinformatical analysis was conducted using disorder prediction software. Disorder prediction methods are designed to help find boundaries of ordered protein domains and regions of disorder, to allow the experimental study of each domain separately (109). Numerous disorder prediction software exists as well as tools that combine results from many different individual methods. Using meta predictions may lead to increased accuracy (109) and the programs metaPrDos and DisMeta were utilized for disorder prediction. Results from the disorder prediction indicate considerable disorder at each terminus of every construct. The disorder prediction of $BcsC^{TPR\ 4\text{-}21}$ presented in Figure 16 illustrates a stretch of 7 residues that appear disordered in addition to the disordered termini. Interestingly, some other constructs

contain this amino acid series yet do not show disorder above the 5% false positive confidence, so this may not be significant as the difference is minimal.

Table 1. ProtParam Protein Features for BcsC Constructs

BcsC TPR Construct Designation	TPR 1-8	TPR 1-11	TPR 1-15	TPR 4-21	TPR 9-21	TPR 12-21
Amino Acids (Start and End)	318 (24-294)	434 (24-410)	582 (24-558)	675 (145-820)	471 (349-820)	361 (459-820)
Molecular Weight	34044.03	47137.53	63926.10	73944.04	52238.73	39858.08
Extinction coefficient (M^{-1} cm^{-1}, at 280 nm measured in water)	21430	41370	69330	90300	67840	53400
Theoretical pI	6.41	7.99	6.67	5.83	5.64	5.41
Instability index	40.31	41.02	44.26	39.81	42.94	39.58
Methionine residues	4	6	8	11	8	6
Cysteine residues	0	0	0	0	0	0
Tryptophan residues	2	4	8	11	8	7
His_6-tag location	N-terminal	N-terminal	N-terminal	C-terminal	C-terminal	C-terminal

*The *E. coli* K12 BcsC amino acid sequence can be found in Appendix Section A1. The values were calculated with the inclusion of His_6-tags.

A) **B)**

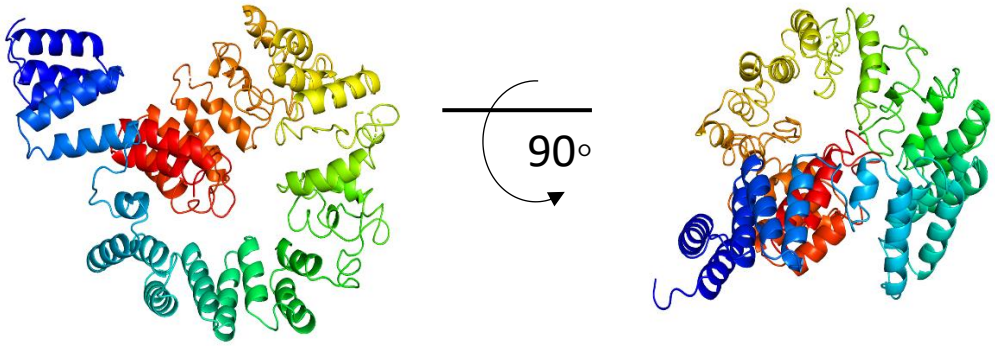


Figure 9. Phyre[2] Predicted Structural Model of BcsC[TPR 1-21]. Phyre[2] model of BcsC[TPR 1-21] (based on PDB: 4BUJ 14% identity) in its native state observed in one superhelical conformation displayed at 2 separate angles (A, & B). Each helix-turn-helix (TPR) is coloured from N-terminal (red) to C-terminal (blue) for distinction. All images were rendered in PyMOL.

A) **B)**

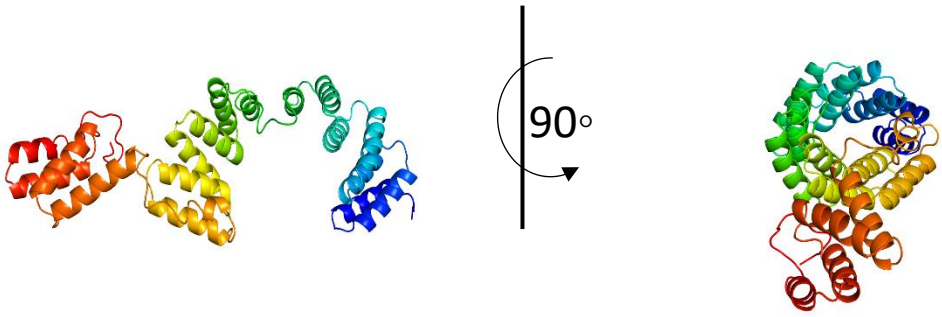


Figure 10. Phyre2 Predicted Structural Model of BcsCTPR [1-8]. Phyre[2] model of BcsC[TPR 1-8] (based on PDB: 4HNX 11% identity) in its native state observed in one superhelical conformation displayed at 2 separate angles (A, & B). Each helix-turn-helix (TPR) is coloured from N-terminal (red) to C-terminal (blue) for distinction. All images were rendered in PyMOL.

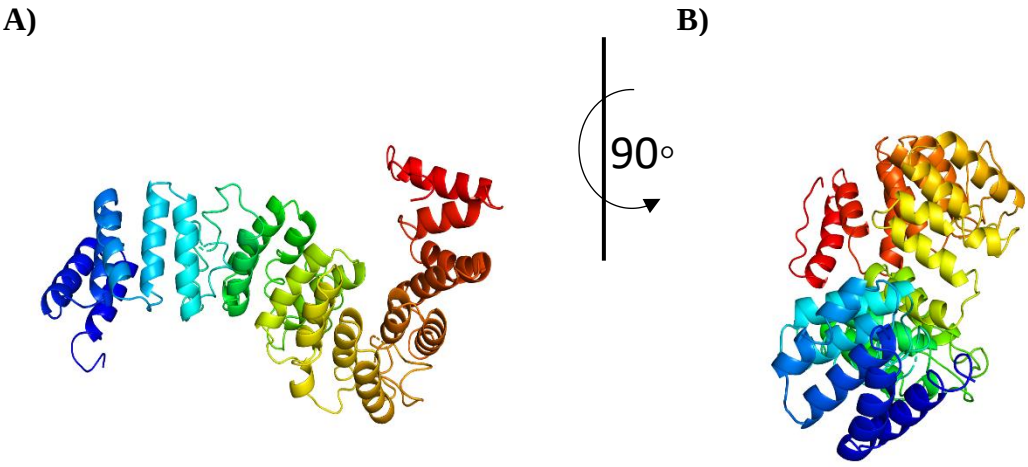


Figure 11. Phyre[2] Predicted Structural Model of BcsC$^{TPR\ 1-11}$. Phyre^2 model of BcsC$^{TPR\ 1-11}$ (based on PDB: 1W3B 13% identity) in its native state observed in one superhelical conformation displayed at 2 separate angles (A, & B). Each helix-turn-helix (TPR) is coloured from N-terminal (red) to C-terminal (blue) for distinction. All images were rendered in PyMOL.

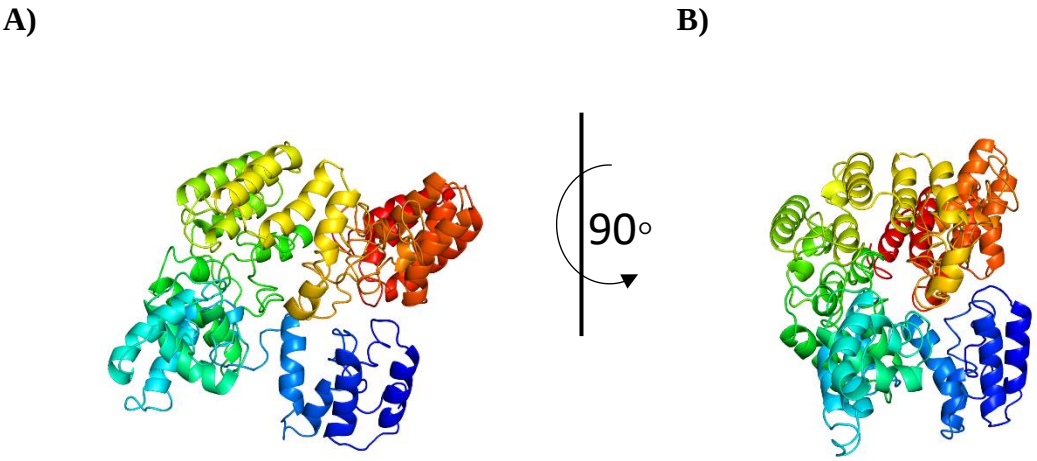


Figure 12. Phyre2 Predicted Structural Model of BcsC$^{TPR\ 1-15}$. Phyre^2 model of BcsC$^{TPR\ 1-15}$ (based on PDB: 5NNR 12% identity) in its native state observed in one superhelical conformation displayed at 2 separate angles (A, & B). Each helix-turn-helix (TPR) is coloured from N-terminal (red) to C-terminal (blue) for distinction. All images were rendered in PyMOL.

A) **B)**

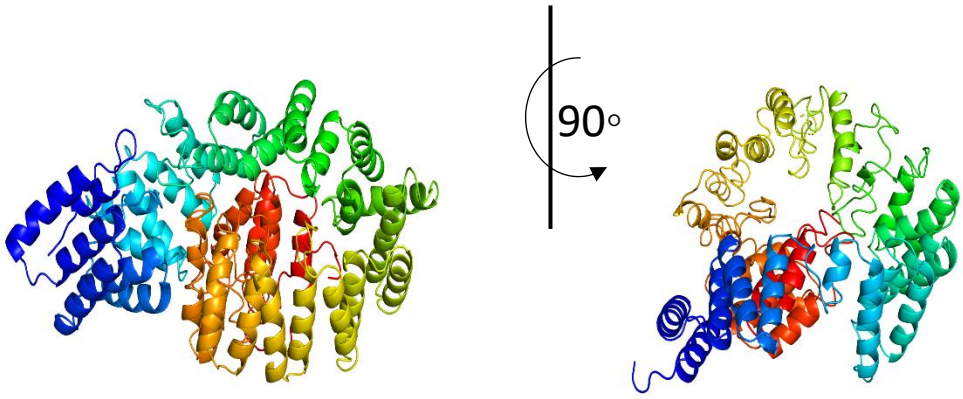


Figure 13. Phyre2 Predicted Structural Model of BcsC$^{TPR\ 4\text{-}21}$. Phyre^{2} model of BcsC$^{TPR\ 4\text{-}21}$ (based on PDB: 4BUJ 14% identity) in its native state observed in one superhelical conformation displayed at 2 separate angles (A, & B). Each helix-turn-helix (TPR) is coloured from N-terminal (red) to C-terminal (blue) for distinction. All images were rendered in PyMOL.

A) **B)**

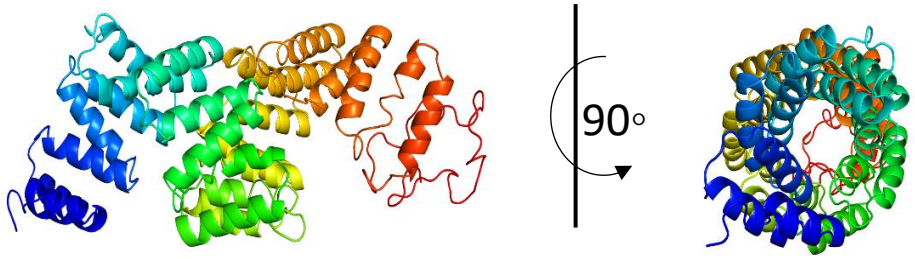


Figure 14. Phyre2 Predicted Structural Model of BcsC$^{TPR\ 9\text{-}21}$. Phyre^{2} model of BcsC$^{TPR\ 9\text{-}21}$ (based on PDB: 4UZY 14% identity) in its native state observed in one superhelical conformation displayed at 2 separate angles (A, & B). Each helix-turn-helix (TPR) is coloured from N-terminal (red) to C-terminal (blue) for distinction. All images were rendered in PyMOL.

A) B)

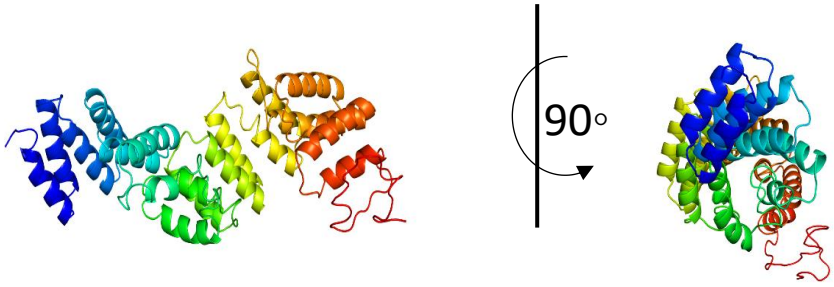


Figure 15. Phyre2 Predicted Structural Model of BcsC$^{TPR\ 12-21}$. Phyre^{2} model of BcsC$^{TPR\ 12-21}$ (based on PDB: 4HNX 11% identity) in its native state observed in one superhelical conformation displayed at 2 separate angles (A, & B). Each helix-turn-helix (TPR) is coloured from N-terminal (red) to C-terminal (blue) for distinction. All images were rendered in PyMOL.

Table 2. BcsCs Percent Identity with AlgK

Protein Name	Sequence Identity %
AlgK	100.00
BcsC$^{TPR\ 1-21}$ (Full Length TPR Domain)	24.46
BcsC$^{TPR\ 4-21}$	24.46
BcsC$^{TPR\ 9-21}$	23.96
BcsC$^{TPR\ 12-21}$	25.93
BcsC$^{TPR\ 1-8}$	24.49
BcsC$^{TPR\ 1-11}$	22.89
BcsC$^{TPR\ 1-15}$	22.51

* Clustal Omega generated percent identity matrix based on amino acid sequences was used to construct this table

Table 3. BcsC Construct Percent Identity to Enterobacter CJF-002 BcsC TPR Domain

Protein Name	Sequence Identity %
Enterobacter CJF-002 BcsC	100.00
BcsC$^{TPR\ 1-21}$ (Full Length TPR Domain)	72.38
BcsC$^{TPR\ 4-21}$	71.32
BcsC$^{TPR\ 9-21}$	72.92
BcsC$^{TPR\ 12-21}$	71.03
BcsC$^{TPR\ 1-8}$	71.75
BcsC$^{TPR\ 1-11}$	73.32
BcsC$^{TPR\ 1-15}$	72.19

* Clustal Omega generated percent identity matrix based on amino acid sequences was used to construct this table

Table 4. Top 5 BLAST Results with PSI-BLAST for BcsC$^{TPR\ 1-21}$

PDB Title	Max Score*	Total Score	Query coverage	% I.D.	PDB I.D.
Crystal structure of the flexible tandem repeat domain of bacterial cellulose synthase subunit C	362	362	42%	72	5XW7_A
Crystal Structure of an 8 Repeat Consensus TPR Superhelix	63.9	208	37%	32	2FO7_A
Design of Stable Alpha-Helical Arrays from An Idealized TPR Motif	57.8	106	28%	31	1NA0_A
Crystal structure of CTPR3Y3	47.8	126	28%	30	2WQH_A
Designed TPR Module (Ctpr390) In Complex with Its Peptide-Ligand (Hsp90 Peptide)	47.4	128	33%	29	3KD7_A

***** Order is ranked by the maximum score and the total score is presented along with the query coverage, percent identity and the PDB identification number.

Table 5. Top 5 Phyre^{2} Predicted Homology Results for BcsC$^{TPR\ 1-21}$

PDB	Coverage	Confidence	% i.d.	PDB Molecule	PDB Title
1. c5nnrD	99%	100.0	12	N-terminal acetyltransferase-like protein	Structure of naa15/naa10 bound to hypk-thb
2. c4bujF	99%	100.0	15	Superkiller protein 3	Crystal structure of the s. cerevisiae ski2-3-8 complex
3. c6c95A	99%	100.0	11	N-alpha-acetyltransferase 15, nata auxiliary	The human nata (naa10/naa15) amino-terminal acetyltransferase complex2 bound to hypk
4. c4hnxA	99%	100.0	10	N-terminal acetyltransferase a complex subunit nat1	The nata acetyltransferase complex bound to ppgpp
5. c4uzyA	98%	100.0	11	Flagellar associated protein	Crystal structure of the Chlamydomonas ift70 and ift52 complex

Table 6. TPRpred Results from BcsC Full Length TPR Domain.

TPR repeat number	Begin	Alignment	Ending Amino Acid Residue Number
1	5	QQQLLEQVRLGEATHREDLVQQSLYRLELIDPNN	38
2	39	PDVVAARFRSLLRQGDIDGAQKQLDRLSQLAPSS	72
3	82	MLLSTPDGRQALQQARLQATTGHAEEAVASYNKL	115
4	124	DIAVEYWSTVAKIPARRGEAINQLKRINADAPGN	157
5	158	TGLQNNLALLLFSSDRRDEGFAVLEQMAKSNAGR	191
6	246	AFRARAQGLAAVDSGMAGKAIPELQQAVRANPKD	279
7	280	SEALGALGQAYSQKGDRANAVANLEKALALDPHS	313
8	328	YWLAIQQGDAALKANNPDRAERLFQQARNVDNTD	361
9	362	SYAVLGLGDVAMARKDYPAAERYYQQTLRMDSGN	395
10	396	TNAVRGLANIYRQQSPEKAEAFIASLSASQRRSI	429
11	438	NDRLAQQAEALENQGKWAQAAALQRQRLALDPGS	471
12	472	VWITYRLSQDLWQAGQRSQADTLMRNLAQQKSND	505
13	506	PEQVYAYGLYLSGHDQDRAALAHINSLPRAQWNS	539
14	544	LVNRLQSDQVLETANRLRESGKEAEAEAMLRQQP	577
15	580	TRIDLTLADWAQQRRDYTAARAAYQNVLTREPAN	613
16	614	ADAILGLTEVDIAAGDKAAARSQLAKLPATDNAS	647
17	648	LNTQRRVALAQAQLGDTAAAQRTFNKLIPQAKSQ	681
18	688	AMVLRDGAKFEAQAGDPTQALETYKDAMVASGVT	721

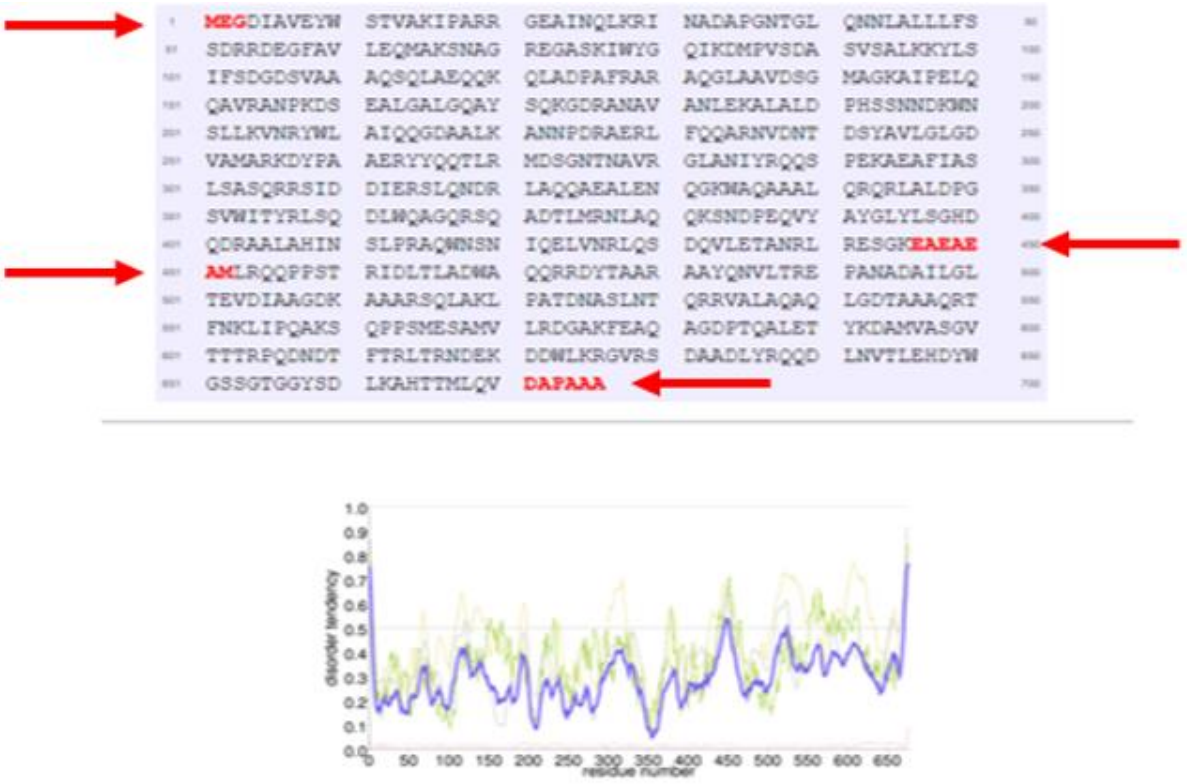


Figure 16. meta PrDos Disorder Prediction Results of BcsC$^{TPR\ 4\text{-}21}$. The top panel contains a numbered amino acid sequence of BcsC$^{TPR\ 4\text{-}21}$ where the red letters indicate potential disordered residues. The bottom graph is the disorder profile plot where the y-axis represents disorder tendency and the x-axis is the BcsC$^{TPR\ 4\text{-}21}$ residue number. The purple line represents the average of multiple disorder prediction programs. The other programs are signified by a mixture of different coloured lines.

4.2 Protein Expression and Purification

4.2.1 Protein Expression

The *E. coli* cells containing the BcsC optimized genes all followed a similar trend during protein expression. Numerous trials of protein expression were conducted in previous experiments over a range of IPTG concentrations and induction temperatures (4°C, 16°C, 22°C and 37°C; this thesis work and Anderson & Weadge, unpublished work). The optimal temperature was found to be 22°C with an IPTG concentration between 0.5 and 1 mM. The duration of the protein expression phase after induction with IPTG was assessed between 4 and

16 hours and negligible differences in downstream protein yields or solubility were noted. This trend is exemplified by comparing BcsC$^{TPR\ 4\text{-}21}$ (Figure 17; ~74.1 kDa), BcsC$^{TPR\ 1\text{-}11}$ (Figure 18; 47.3 kDa), and BcsC$^{TPR\ 1\text{-}15}$ (Figure 19; 64 kDa) expressed fractions in the 4 h post-induction (lane 3 in each of the figures) and 16 h post-induction (lane 4 in each of the figures) samples. You will have to note, that although lanes for expression of BcsC$^{TPR\ 1\text{-}8}$, BcsC$^{TPR\ 9\text{-}21}$, and BcsC$^{TPR\ 12\text{-}21}$ are not depicted in Figures 20, 21, and 22 respectively, they followed a similar trend to BcsC$^{TPR\ 1\text{-}15}$ (Figure 19). The only exception to this trend is noted with the respective protein band for BcsC$^{TPR\ 4\text{-}21}$ at the 16 h mark being less than the 4 h mark. Previous studies have identified that in some cases, less soluble protein was produced due to the formation of inclusion bodies, through aggregation, within *E.coli* expression systems at molecular weights of over ~60 kDa (140). In these cases, a longer expression may allow protein to degrade, thus, lowering the amount and the quality of the soluble protein yield. Despite the degradation and/or aggregation that may be associated with this construct (if any), the total amount of soluble protein gained from the extra hours of expression was deemed to out-weigh any amount lost.

The largest protein yields after secondary purification were with BcsC$^{TPR\ 1\text{-}11}$ and BcsC$^{TPR\ 12\text{-}21}$ (Table 7). Proteins over ~60 kDa may exhibit less successful expression of soluble protein in *E. coli* (140) and BcsC$^{TPR\ 4\text{-}21}$ and BcsC$^{TPR\ 1\text{-}15}$ are both over 60 kDa, while BcsC$^{TPR\ 9\text{-}21}$ (~52.4) is fairly close to this threshold. While size may play a factor, it is unclear precisely why protein expression was different among constructs and could also entail better expression or binding to the purification columns under the conditions used. Regardless, most of the constructs yielded protein amounts that were amenable to downstream analyses, so once these concentrations were achieved with a standard protocol, further optimization was not pursued.

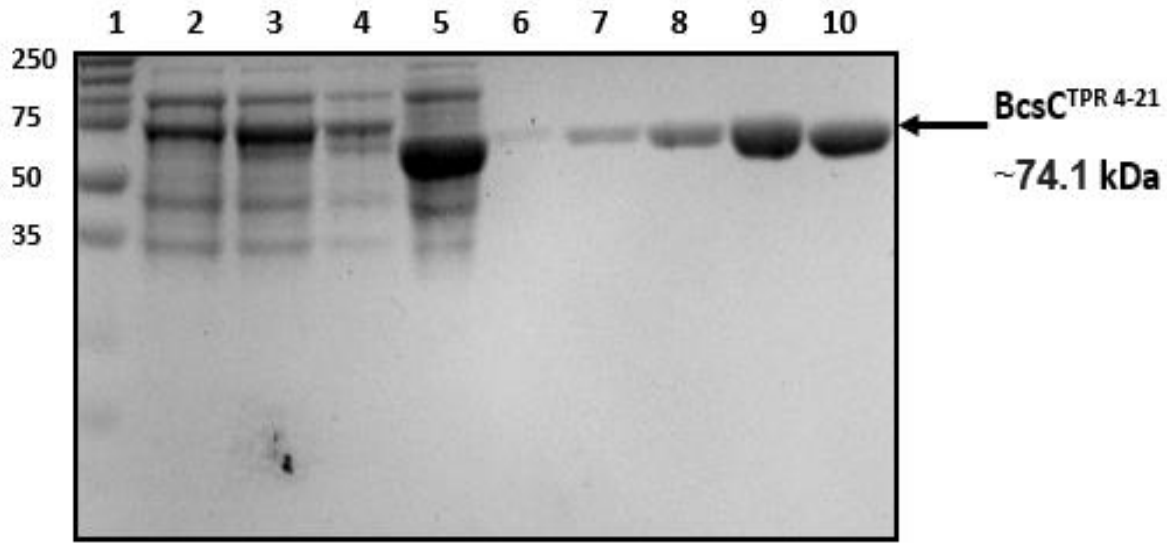


Figure 17. SDS-PAGE Protein Expression and Purification of BcsC$^{TPR\ 4\text{-}21}$. SDS-PAGE analysis (12% (v/v)) of BcsC$^{TPR\ 4\text{-}21}$ expression of 1 L cultures incubated at 22°C and 200 rpm shaking. Lane 1, molecular-weight markers (kDa); lane 2, before induction sample; lane 3, expression time 4 h sample; lane 4 expression time 16 h; lane 5, Ni-NTA column unbound lysate; lane 6, wash with lysis buffer (no imidazole); lane 7, wash 1 (20 mM imidazole); lane 8, wash 2 (40 mM imidazole); lane 9, elution buffer (250 mM imidazole); lane 10, elution fraction from anion exchange chromatography column.

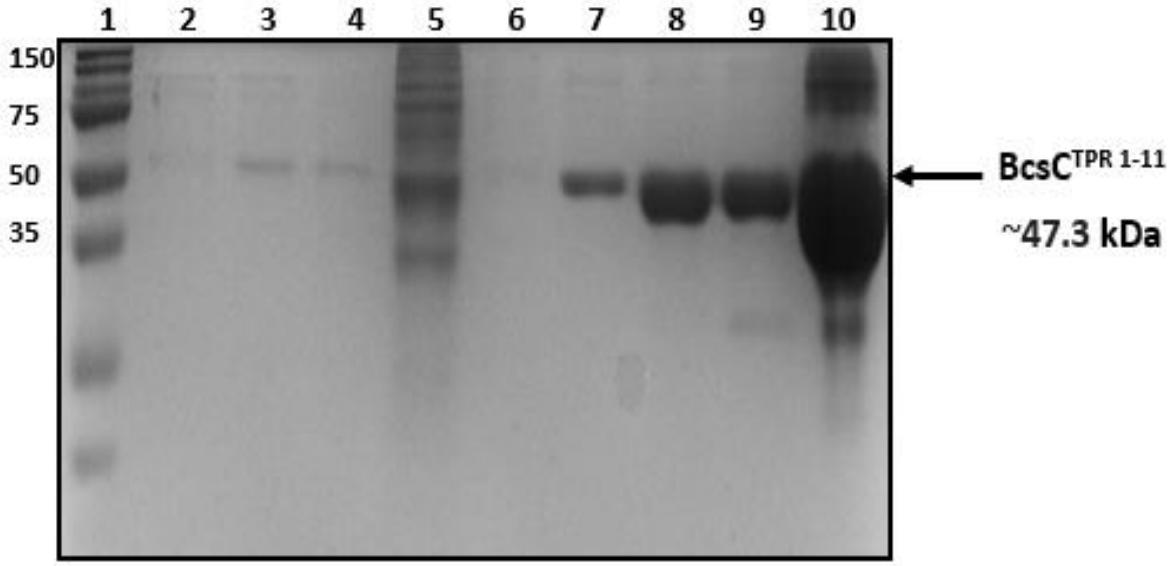


Figure 18. SDS-PAGE Protein Expression and Purification of BcsC$^{TPR\ 1-11}$. SDS-PAGE analysis (12% (v/v)) of BcsC$^{TPR\ 1-11}$ expression of 1 L cultures incubated at 22°C and 200 rpm shaking, Ni-NTA purification fractions, and a concentrated ion exchange chromatography sample. Lane 1, molecular-weight markers (kDa); lane 2, expression induction time 0 sample; lane 3, expression time 4 h sample; lane 4 expression time 16 h; lane 5, Ni-NTA column unbound lysate; lane 6, wash with lysis buffer (no imidazole); lane 7, wash 1 (20 mM imidazole); lane 8, wash 2 (40 mM imidazole); lane 9, elution buffer (250 mM imidazole); lane 10, sample of concentrated fractions pooled following cation exchange chromatography.

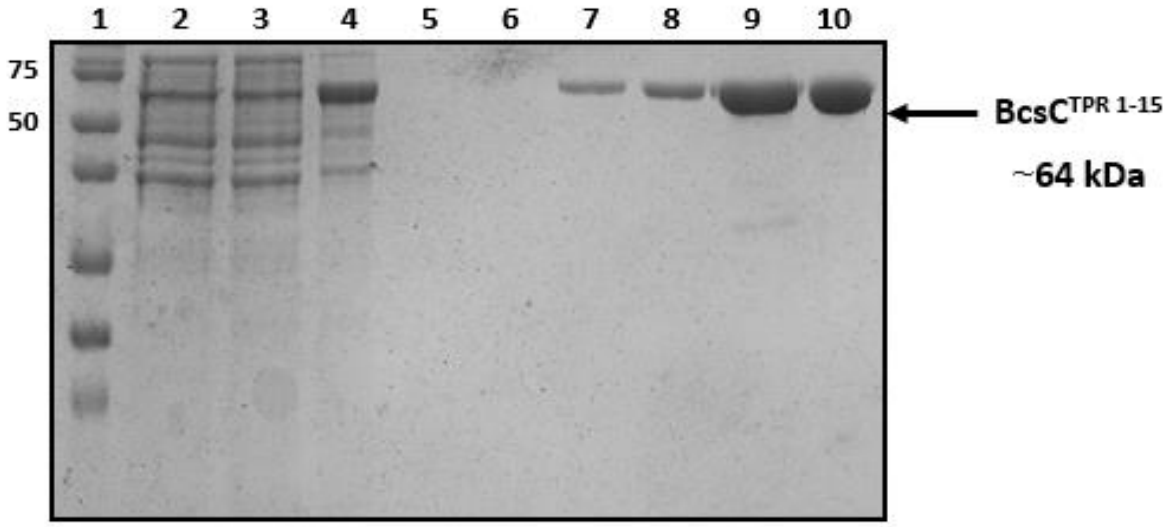


Figure 19. SDS-PAGE Protein Expression and Purification of BcsC$^{TPR\ 1-15}$. SDS-PAGE analysis (12% (v/v)) of BcsC$^{TPR\ 4-21}$ expression of 1 L cultures incubated at 22°C and 200 rpm shaking, Ni-NTA purification fractions, and anion exchange flow through. Lane 1, molecular-weight markers (kDa); lane 2, expression induction time 0 sample; lane 3, expression time 4 h sample; lane 4 expression time 16 h; lane 5, Ni-NTA column unbound lysate; lane 6, wash with lysis buffer (no imidazole); lane 7, wash 1 (20 mM imidazole); lane 8, wash 2 (40 mM imidazole); lane 9, elution buffer (250 mM imidazole); lane 10, flow through from anion exchange chromatography column.

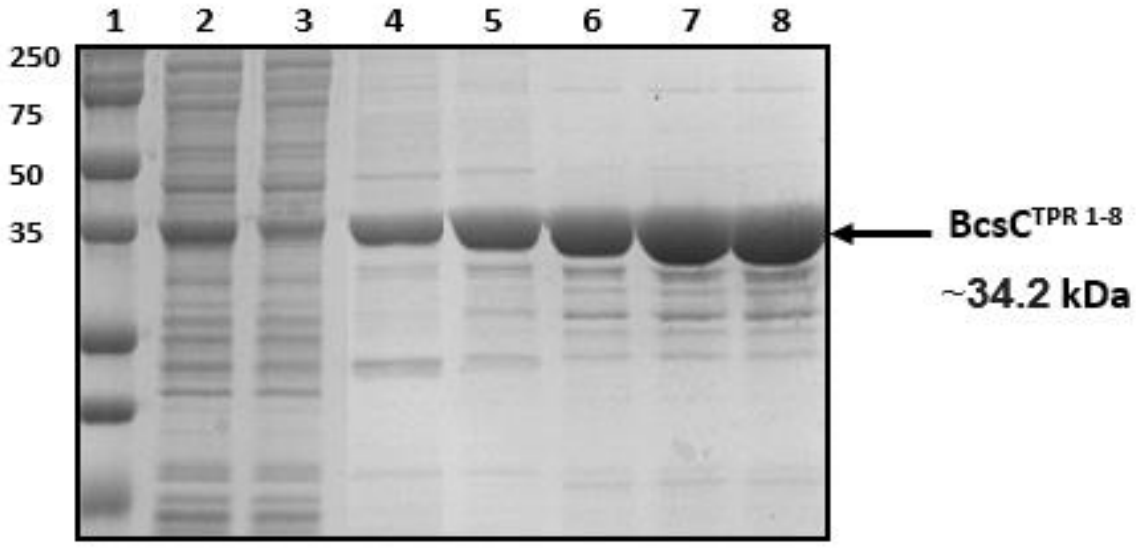


Figure 20. SDS-PAGE Purification of BcsC$^{TPR\ 1-8}$. SDS-PAGE analysis (12% (v/v)) of BcsC$^{TPR\ 1-8}$ Ni-NTA purification fractions. Lane 1, molecular-weight markers (kDa); lane 2, Ni-NTA column unbound lysate; lane 3, wash with lysis buffer (no imidazole); lane 4, wash 1 (20 mM imidazole); lane 5, wash 2 (40 mM imidazole); lanes 6-8, elution buffer (250 mM imidazole).

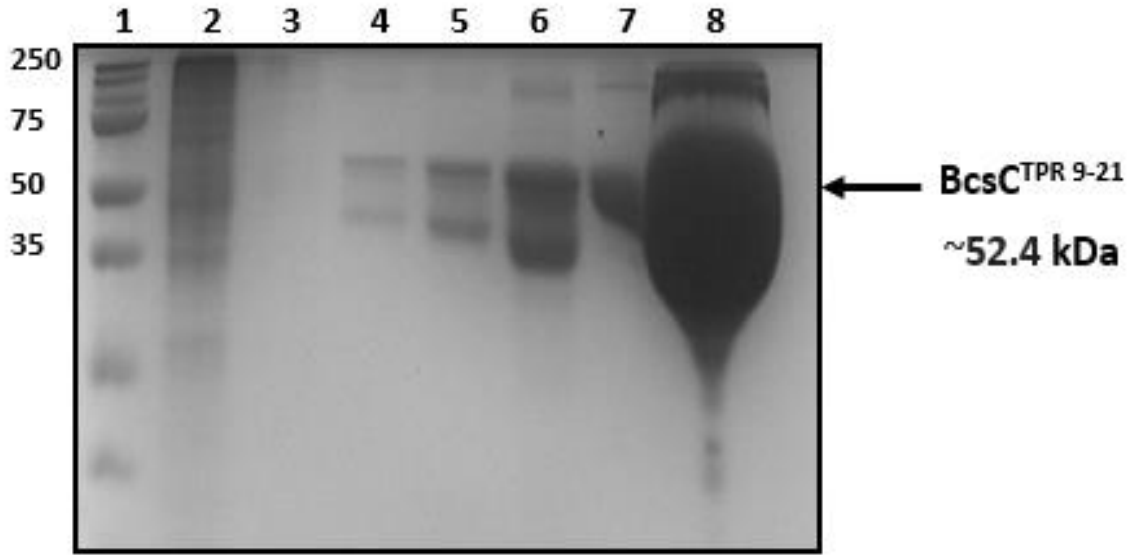


Figure 21. SDS-PAGE Purification of BcsC$^{TPR\ 9-21}$. SDS-PAGE analysis (12% (v/v)) of BcsC$^{TPR\ 9-21}$ Ni-NTA purification fractions, an anion exchange chromatography fraction and a concentrated sample of anion fractions. Lane 1, molecular-weight markers (kDa); lane 2, Ni-NTA column unbound lysate; lane 3, wash with lysis buffer (no imidazole); lane 4, wash 1 (20 mM imidazole); lane 5, wash 2 (40 mM imidazole); lanes 6, elution buffer (250 mM imidazole); lane 7, anion exchange elution fraction; lane 8, concentrated sample of anion exchange fractions.

4.2.2 Immobilized Metal Affinity Chromatography

After protein expression, IMAC purification was utilized to isolate the his_6-tagged recombinant proteins for all constructs. The only construct that was not be purified past IMAC was BcsC$^{TPR\ 1-8}$ due to greater difficulties with degradation compared to most other constructs. This construct was not pursued further as this TPR region was already covered with other protein constructs (BcsC$^{TPR\ 1-11}$, BcsC$^{TPR\ 1-15}$) that behaved noticeably better during purification. Purification of all constructs followed a common route that involved loading lysate onto the column in the absence of imidazole and then gradually increasing the presence of this molecule through the washing steps (up to 40mM) to remove background contaminants and then finally eluting in 250 mM imidazole. SDS PAGE images of BcsC$^{TPR\ 4-21}$, BcsC$^{TPR\ 1-11}$, and BcsC$^{TPR\ 1-15}$ (Figures 17, 18, and 19; lanes 5-9) as well as BcsC$^{TPR\ 1-8}$, BcsC$^{TPR\ 9-21}$ (Figures 20 and 21; lanes

5-8) and BcsC$^{TPR\ 12-21}$ (Figures 22; lanes 3-10) contain typical IMAC purification samples. Although some protein was lost in the wash steps the majority of the protein was contained in the elutions across all protein constructs as indicated through strong bands focused around the appropriate molecular weights (BcsC$^{TPR\ 4-21}$ 74.1 kDa~, BcsC$^{TPR\ 1-11}$ ~43.3 kDa BcsC$^{TPR\ 1-15}$ ~64 kDa, BcsC$^{TPR\ 1-8}$ ~34.2 kDa, BcsC$^{TPR\ 9-21}$ ~52.4 kDa, and BcsC$^{TPR\ 12-21}$ ~40 kDa, respectively). A significant amount of contaminating protein (as seen by additional bands on the SDS-PAGE gels) was noted for BcsC$^{TPR\ 1-8}$ (Figure 20; lanes 6-8) and BcsC$^{TPR\ 9-21}$(Figure 21; lane 6), thereby, prompting a definite need for secondary purification with these two constructs. In contrast, the BcsC$^{TPR\ 12-21}$ elutions (Figure 22; Lanes 6-10) contained significantly less contamination, but secondary purification was still performed (outlined below) to further remove contaminants even for this construct on a routine basis so that downstream analyses would have the purest protein possible.

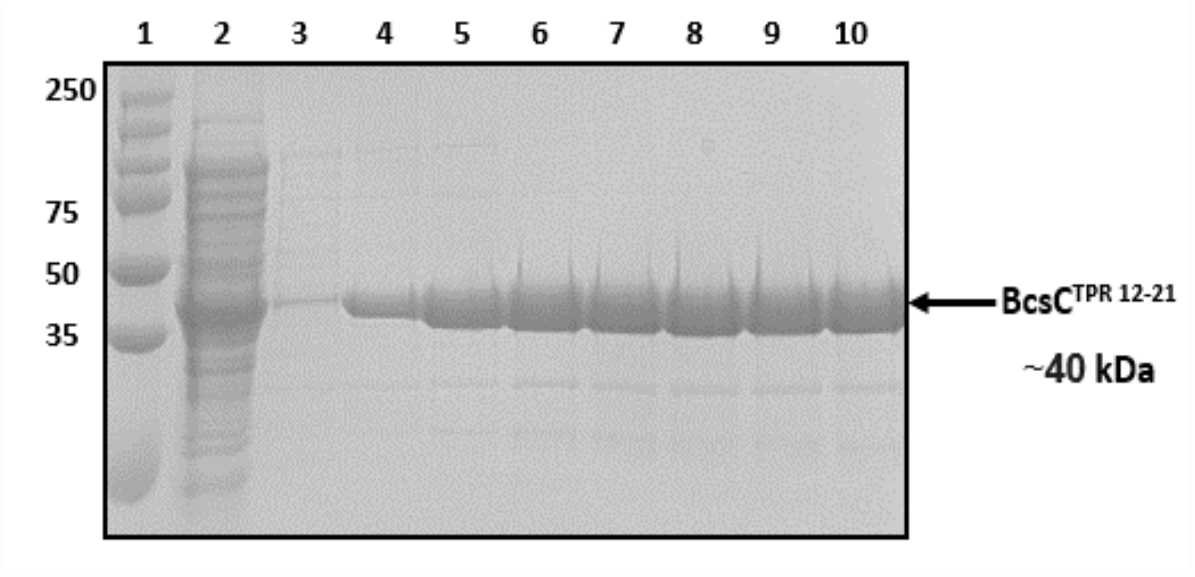


Figure 22. SDS-PAGE Purification of BcsC$^{TPR\ 12-21}$. SDS-PAGE analysis (12% (v/v)) of BcsC$^{TPR\ 12-21}$ Ni-NTA purification fractions. Lane 1, molecular-weight markers (kDa); lane 2, Ni-NTA column unbound lysate; lane 3, wash with lysis buffer (no imidazole); lane 4, wash 1 (20 mM imidazole); lane 5, wash 2 (40 mM imidazole); lanes 6-10, elution buffer (250 mM imidazole).

4.2.3 Ion Exchange Chromatography

Following pooling of elutions, dialysis to remove salts and concentration of protein constructs, they were subjected to ion exchange chromatograph. Ion exchange profiles were similar for all constructs tested and a typical profile is displayed in Figure 23 for BcsC$^{TPR\ 9\text{-}21}$. The ion exchange profile is signified by a large blue "bump" that represents the protein being passed over the ion column numerous times to facilitate ample time for binding of BcsC to the ion column. This is followed by a large dip in the blue line that represents buffer containing no salt removing protein that did not bind to the column (some contaminations are lost). Elution of target proteins from the column were observed as a second peak in the profile and the apex for this peak for all constructs tested is presented in Table 8 as a comparison of elution times/volumes. As a companion to these profiles to judge purity and the presence of the desired protein construct, SDS-PAGE analysis of the elution fractions was also performed (example elutions for BcsC$^{TPR\ 12\text{-}21}$ and BcsC$^{TPR\ 9\text{-}21}$depicted in Figure 24). A comparison of the ion exchange elutions to the earlier IMAC elutions (*ie.* Figure X and Y for BcsC$^{TPR\ 12\text{-}21}$ and BcsC$^{\mathbf{TPR\ 9\text{-}21}}$, respectively) clearly indicated that ion exchange chromatography greatly increased the protein purity. This was true for all constructs where the purity of BcsC$^{TPR\ 4\text{-}21}$ (Figure 17; lane 10) BcsC$^{TPR\ 1\text{-}11}$ (Figure 18; lane 10) and BcsC$^{TPR\ 1\text{-}15}$ (Figure 19; lane 10) was better compared to the previous lane (IMAC elution) in all of the respective figures.

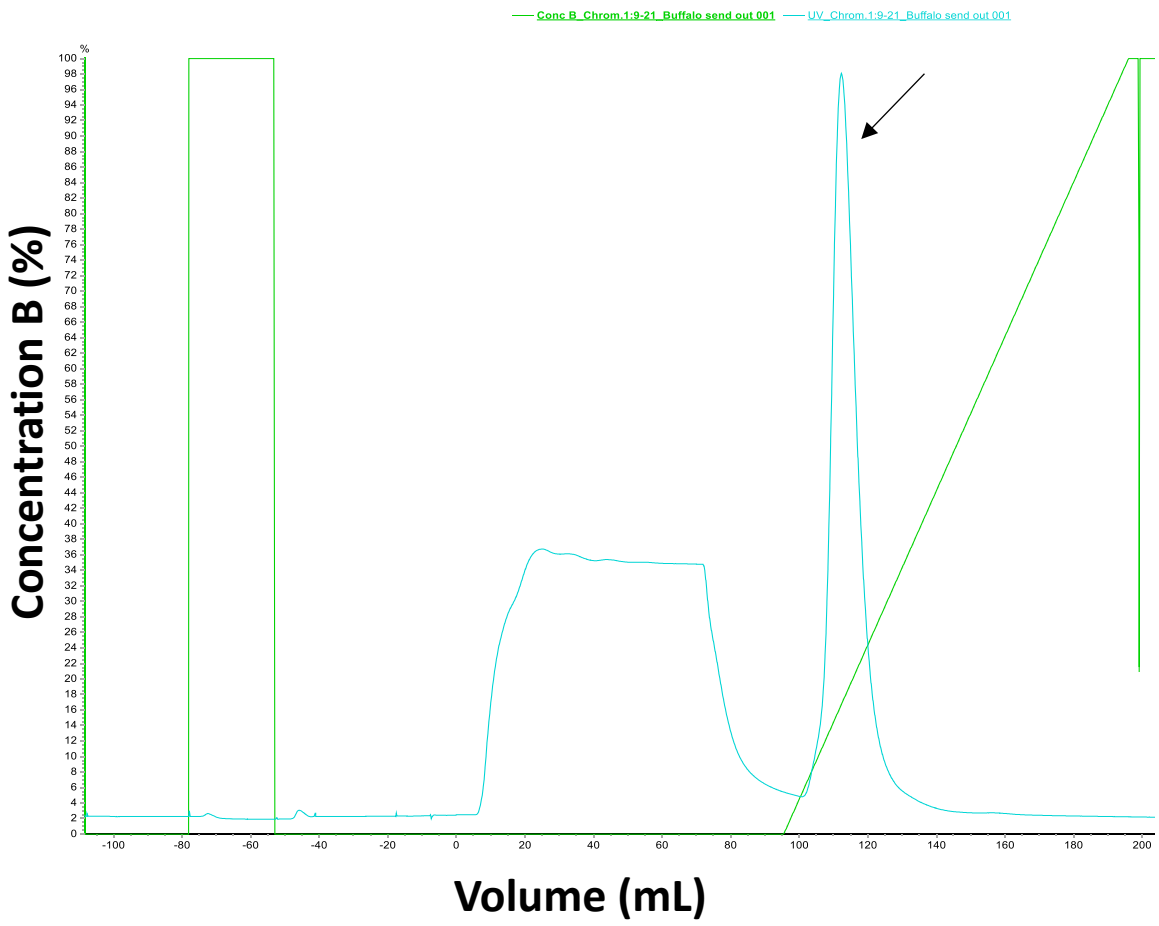


Figure 23. Example Ion Exchange Purification Profile of BcsC$^{TPR\ 9\text{-}21}$. Monitoring of elution from the anion exchange column for the BcsC$^{TPR\ 9\text{-}21}$ construct was accomplished by recording the absorbance at 305 nm (Blue line) in conjunction with elution using an increasing concentration of NaCl (represented by the green line). The black arrow denotes the elution peak for BcsC$^{TPR\ 9\text{-}21}$ that was collected and further analyzed via SDS PAGE. This chromatogram is a typical profile for each of the BcsC constructs tested through Ion Exchange Chromatography.

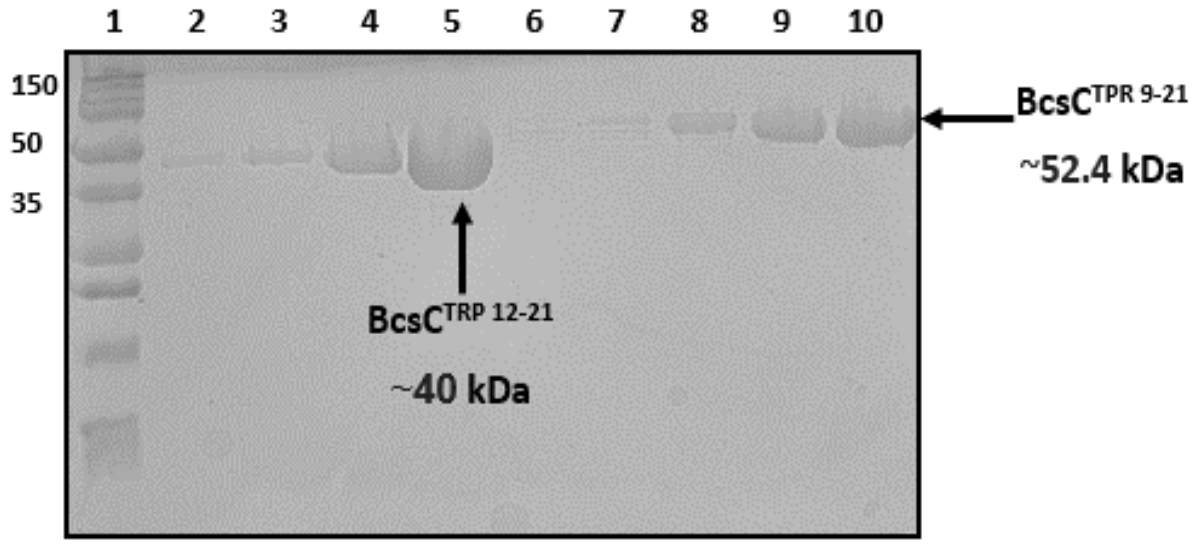


Figure 24. Anion Exchange Purification of BcsC$^{TPR\ 12-21}$ and BcsC$^{TPR\ 9-21}$.SDS-PAGE analysis (12% (v/v)) of BcsC$^{TPR\ 12-21}$ and BcsC$^{TPR\ 9-21}$ anion exchange chromatography fractions. Lane 1, molecular-weight markers (kDa); lanes 2-5, BcsC$^{TPR\ 12-21}$ anion exchange elution fractions; lanes 6-10, BcsC$^{TPR\ 9-21}$ anion exchange elution fractions.

Table 7. BcsC Protein Construct Yields

Protein Construct	Total protein yield (per litre of culture purified)*
BcsC$^{TPR\ 1-8}$	Not Determined
BcsC$^{TPR\ 1-11}$	~6-7 mg/L
BcsC$^{TPR\ 1-15}$	~3-4 mg/L
BcsC$^{TPR\ 4-21}$	~2.5-3 mg/L
BcsC$^{TPR\ 9-21}$	~2.5-3 mg/L
BcsC$^{TPR\ 12-21}$	~6-7 mg/L

*Protein yield after secondary (anion) purification for each BcsC was tabulated across at multiple purifications.

Table 8. Ion Exchange Indicators for BcsC Protein Constructs

Protein Construct	mAU range*	Salt percentage beginning**	Salt Percentage Ending**
BcsC$^{TPR\ 1-11}$	700-1400	19	31
BcsC$^{TPR\ 1-15}$	480-700	5	24
BcsC$^{TPR\ 4-21}$	150+	8	25
BcsC$^{TPR\ 9-21}$	250-350	8	26
BcsC$^{TPR\ 12-21}$	700-1400	9	24
BcsC$^{TPR\ 1-15}$SeMet	300-350	5	20

***** milli-Absorbance Units (mAU) at 305 nm

** indicative of the salt percentage at which the proteins begin to and finishes eluting as judged by the elution profile

4.3 Dynamic Light Scattering (DLS)

Prior to crystal screening attempts, DLS was conducted with fresh protein from the ion exchange columns in order to assess protein stability and polydispersity. Athough, a variety of diluted concentrations of protein were used for analysis, it was found that the elutions coming directly from the ion exchange column yielded the best results for all constructs. Early DLS data samples (Figure 25) indicated that either the concentrations tested were too high or that the protein was polydisperse. For example, analysis of $Bcsc^{TPR\ 1\text{-}15}$ either contained two distinct species (Figure 25A), multiple aggregates (Figure 25B) and/or increasing amounts of a larger polydispersity (Figure 25C). After protein purification was refined to include the use of protease inhibitors and by decreasing the amount of time from cell lysis to downstream applications lower polydispersity levels were observed. For example, ample DLS data from $BcsC^{TPR\ 12\text{-}21}$ (Figure 26; Panel A) and $BcsC^{TPR\ 1\text{-}15}$ (Figure 26; Panel B) indicate monodisperse samples of one species. This data represents 15 image acquisitions at 5 s intervals (averaged) and numerous samples from these two constructs that routinely achieved polydispersity levels of below 20%. These protein constructs were then subjected to protein crystallization attempts.

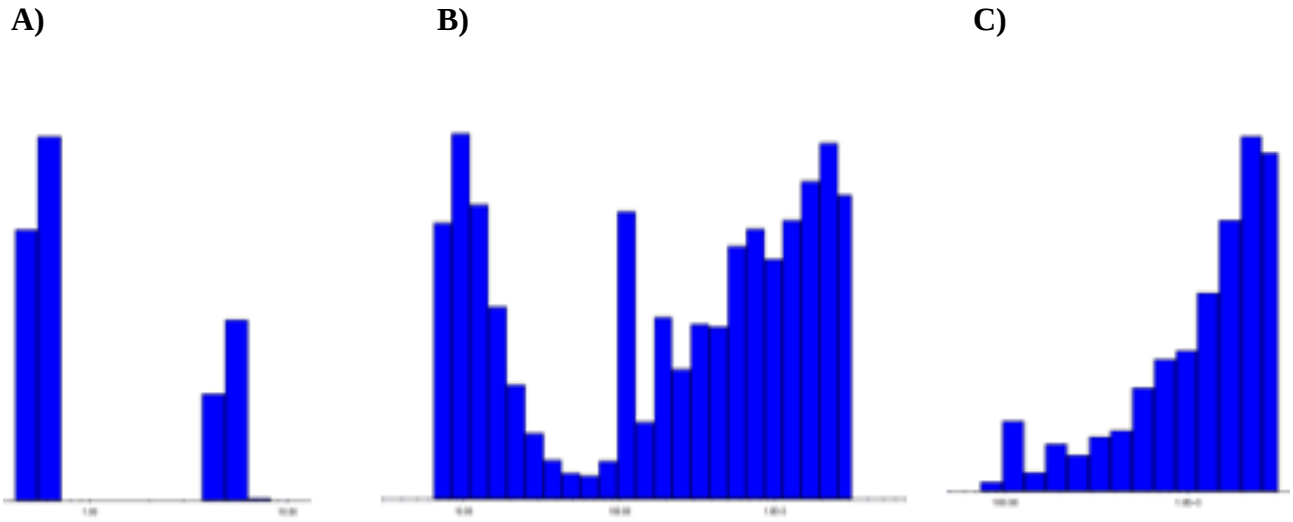
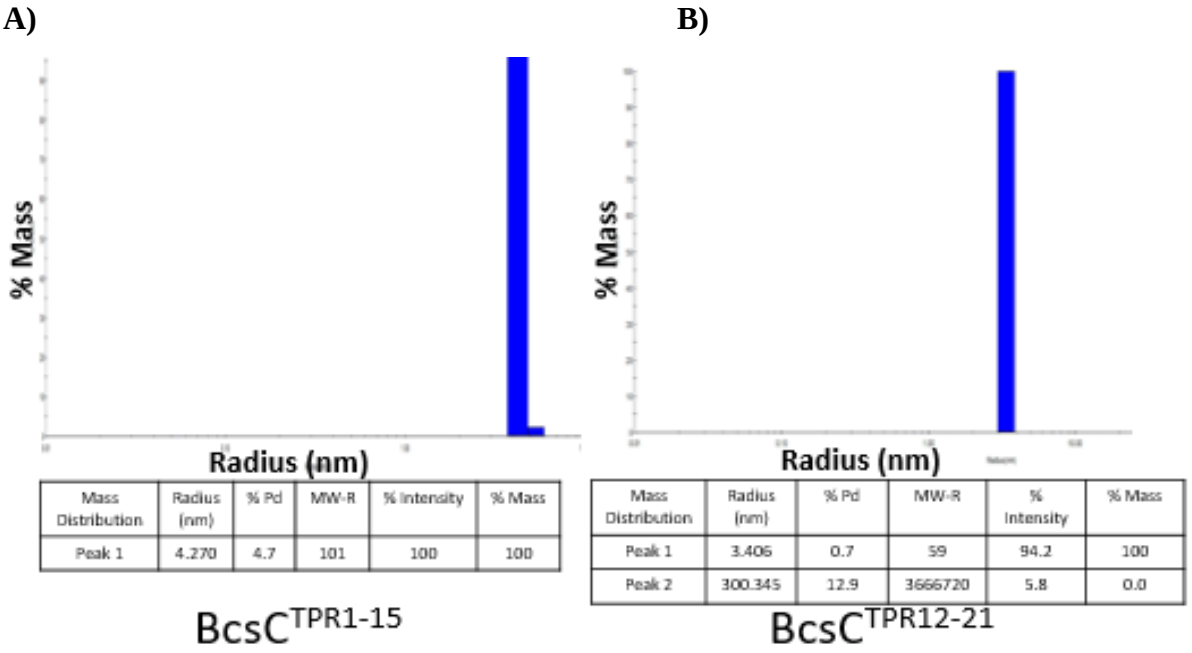


Figure 25. Preliminary DLS Data from BcsC$^{TPR\ 1\text{-}15}$.

Three separate DLS profiles containing high levels of polydispersity. A) Appears to contain two distinct species B) Appears to contain multiple aggregates C) Contains increasing amounts of a larger polydispersity.



Mass Distribution	Radius (nm)	% Pd	MW-R	% Intensity	% Mass
Peak 1	4.270	4.7	101	100	100

BcsC$^{TPR1\text{-}15}$

Mass Distribution	Radius (nm)	% Pd	MW-R	% Intensity	% Mass
Peak 1	3.406	0.7	59	94.2	100
Peak 2	300.345	12.9	3666720	5.8	0.0

BcsC$^{TPR12\text{-}21}$

Figure 26. DLS Dispersity Chart for (A)BcsC$^{TPR\ 1\text{-}15}$(B) BcsC$^{TPR\ 12\text{-}21}$.

Dynamic Light Scattering data containing 15 acquisitions. A) BcsC$^{TPR\ 1\text{-}15}$ protein data indicates an extremely low percentage of polydispersity B) BcsC$^{TPR\ 12\text{-}21}$ protein data indicates nearly 100% mono-dispersity.

4.4 Protein Crystallization

4.4.1 Crystal Screening Trials

Protein crystallization is usually most successful with the highest purity of protein accessible. In most cases that means multiple purification steps, as was the case with BcsC protein constructs. In addition, protein crystallization usually requires additives and precipitating agents to achieve nucleation points in conjunction with the correct ratio of the protein to assorted conditions. Thus, initial crystal trials employ surveying many crystal conditions that cover crystal space that has been successful in the past, along with a series of conditions that explore common buffers, precipitants and salts that researchers use in their purification steps.

For each of the BcsC constructs that yielded pure protein, multiple crystal screens (MCSG 1-4, Top 96, Pure PEG, PACT Premier, JCSG Plus, Morpheus) in conjunction with multiple protein concentrations (13, 15, 20, 22, 26, 29, 30, 32, 35, 40 mg/mL) were surveyed (Table 9). From this research, two constructs, $BcsC^{TPR\ 1\text{-}15}$ and $BcsC^{TPR\ 12\text{-}21}$, resulted in high levels of purity (see Figure 19; lane 10 and Figure 24; lane 5, respectively) early on and, therefore, crystal screening of these constructs was more comprehensive than the other constructs (a conservative estimate of 9,000 conditions each were analyzed). The various conditions that elicited crystal hits from screening these two constructs are presented in Table 10 and Table 11. $BcsC^{TPR\ 1\text{-}11}$ was also purified to a high level through ion exchange (see Figure 18; lane 10) and was subjected to numerous crystal trials (approximately 3000 different conditions in-house) but did not yield crystal hits in any of the conditions surveyed. To accelerate and broaden our screening conditions, the $BcsC^{TPR\ 1\text{-}11}$, $BcsC^{TPR\ 9\text{-}21}$, and $BcsC^{TPR\ 12\text{-}21}$ purified samples were also sent to HWI for their analysis in 1,536-well microassay plates. Screening with HWI was completed twice, where the first shipment included protein that was shipped 16 h

following ion exchange purification (storage and shipment of this protein was held at 4°C until crystal trays were set up). Although there appeared to be crystal growth (Figure 27 A and B), later examination of the SONICC and ultraviolet light images (Figure 27 C and D respectively), conducted at the termination of the experiment, indicated that the crystals were salt in composition (*ie.* protein crystals should fluoresce under UV light due to the presence of tryptophan residues in each protein construct (see Table 1)). HWI screening with the same protein constructs ($BcsC^{TPR\ 1-11}$, $BcsC^{TPR\ 9-21}$, and $BcsC^{TPR\ 12-21}$) was repeated a second time, but this trial differed in that samples were frozen at -80°C directly following purification and shipped to the facility on dry ice. Unfortunately, despite the altered preparation, only salt crystals were once again observed following careful examination of the brightfield microscopy, TPE-UVF, and SONICC images.

Table 9. Protein Concentrations Screened for Each Construct

Protein Construct	Concentrations Screened (mg/mL)	Crystal Screens Used	Total Number of Conditions Screened
$BcsC^{TPR\ 1-8}$	None	None	0
$BcsC^{TPR\ 1-11}$	20, 26, 35, 40	MCSG 1, 2, 3, 4, Morpheus, Pact Premier (BN-156-1-36), JCSG-Plus	~3000 + 1536 (HWI) = ~4500
$BcsC^{TPR\ 1-15}$	13, 15, 20, 24, 26, 29, 32, 35	Top 96, MCSG 1, 2, 3, 4, Morpheus, Pact Premier (BN-156-1-36), JCSG-Plus, Pure PEG	~ 9000
$BcsC^{TPR\ 4-21}$	None	None	0
$BcsC^{TPR\ 9-21}$	None	None	1536 (HWI)
$BcsC^{TPR\ 12-21}$	15, 20, 24, 26, 29, 30, 32, 35	Top 96, MCSG 1, 2, 3, 4, Morpheus, Pact Premier (BN-156-1-36), JCSG-Plus, Pure PEG	~ 9000 + 1536 (HWI) = ~10500

Table 10. Summary of BcsC$^{TPR\ 1-15}$ Crystal Hits

Crystal Screen**	Condition	Concentration and Appearance
MCSG-3 G12 (#84)	0.1 M sodium acetate pH 4.6, 2 M sodium formate	Numerous concentrations Spherulites
Morpheus G3 (#23)	10% (w/v) PEG 4000, 20% glycerol, 0.2 M mixture of carboxylic acids, 0.1 M MES/Imidazole	35 mg/mL Microcrystals
MCSG-2 C8 (#59)	0.1 Bis-Tris Propane: NaOH pH 7.0, 1.8 M MgSulfate	35 mg/mL* Micro-rod
MCSG-2 D8 (#60)	0.2 M K Nitrate pH 6.9, 20% (w/v) PEG 3350	35mg/mL* Micro-rod
MCSG-1 A1 (#1)	0.1 M Hepes: NaOH pH 7.5, 20% (w/v) PEG 8000	13 mg/mL Microcrystal
PACT Premier BN 156-1-36) A10 (#73)	0.2 M MgCl hexahydrate, 0.1 M Sodium Acetate pH 5.0, 20% (w/v) PEG 6000	40 mg/mL* Spherulites-small

*Protein was incubated with D-glucose for a minimum of 1 hour prior to preparation

**The numbers in brackets denote the numerical designation of the condition

Table 11. Summary of BcsC$^{TPR\ 12-21}$ Crystal Hits

Crystal Screen*	Condition	Concentration and Appearance
Top 96 A1 (#1)	0.2 M MgCl, 0.1 M Tris-HCl pH 8.5, 30% (w/v) PEG 4000	20 mg/mL Rod-small
JCSG-Plus D12 (#92)	0.04 M Potassium phosphate monobasic, 16 % w/v PEG 8000, 20 % v/v Glycerol	29 mg/mL Microcrystal
Morpheus H1 (#85)	0.1 M Imidazole; MES pH 6.5, 50% (v/v) precipitant mixture, 0.1 M amino acid mixture	29 mg/mL Microcrystal
MCSG-2 F6 (#66)	0.2 M sodium malonate pH 7, 20% (w/v) PEG 3350	30 mg/mL Cubic-small

*The numbers in brackets denote the numerical designation of the condition

A

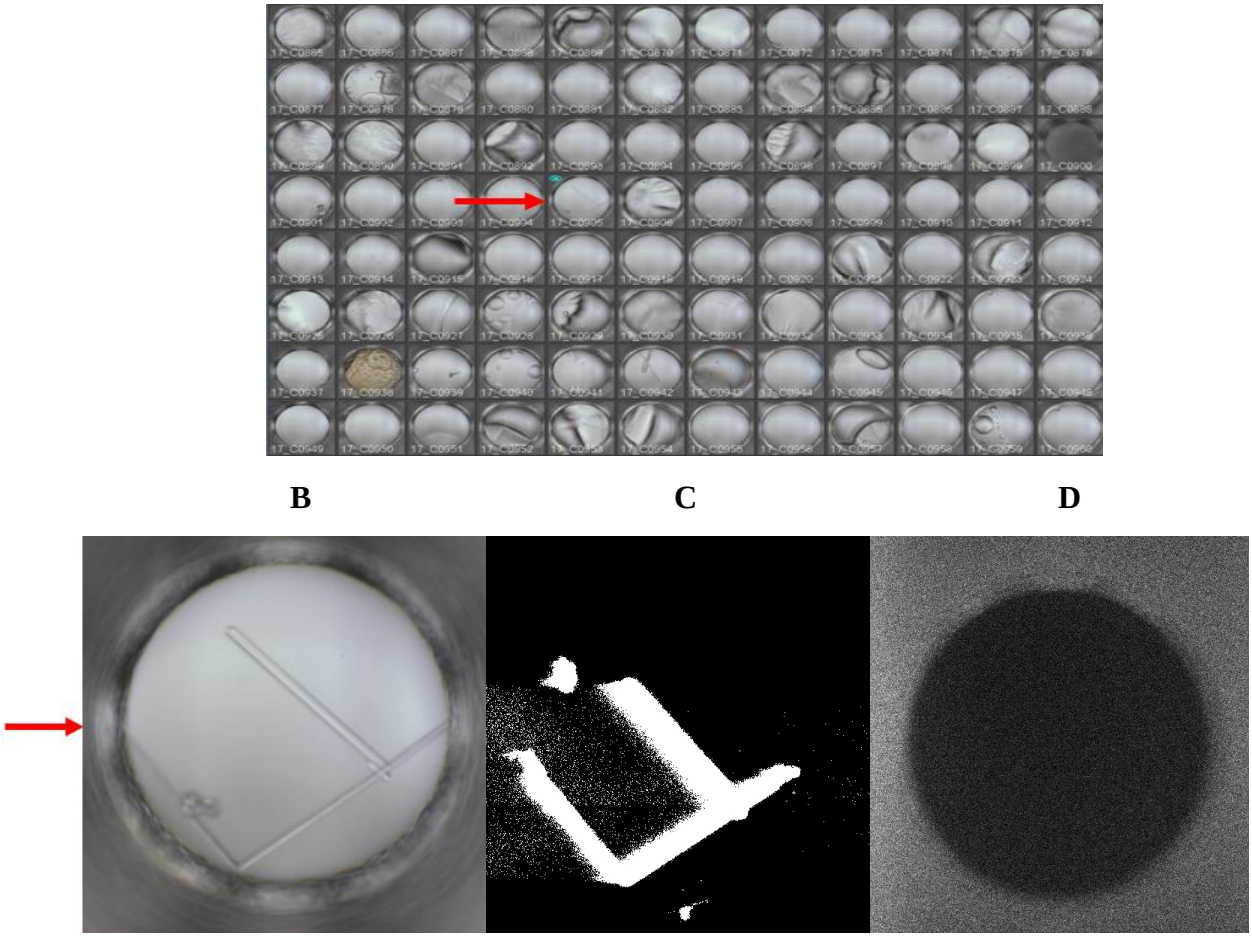


Figure 27. Representative Sample of BcsC$^{TPR\ 12-21}$ High-Throughput Images from HWI.

A) Sample view of 96 wells (with well #905 indicated by the red arrow) from database containing Hauptman-Woodward High-Throughput screening images. B) Sample microscopy image of crystals from well #905 (red arrow). C) SONICC image of well #905 indicates the presence of a crystalline object. D) UV image of 905 illustrates the absence of protein crystals. Microscopy, SONICC, and UV were used in combination to ensure accurate conclusions.

4.4.2 Crystal Expansion Plate Trials

Following initial crystal screening, conditions to recreate successful crystal events in larger plate formats through manual setup were attempted. The first was with BcsC$^{TPR\ 1-15}$ in the condition 0.1 M Hepes: NaOH pH 7.5, 20% (w/v) PEG 8000 and yielded microcrystals surrounded by precipitate. Unfortunately, increasing the concentration did not cause the crystals

to grow large enough for looping. Next the $BcsC^{TPR\ 1\text{-}15}$ condition of 10% (w/v) PEG 4000, 20% glycerol (v/v), 0.2 M mixture of carboxylic acids, 0.1 M MES/Imidazole was attempted in sitting drop 96-well format, but crystal growth did not occur. There were no further attempts as the microcrystals were a lower priority than other larger crystals and the buffer contents of the condition were difficult to recreate without using the remaining buffer within the crystal kit. As IZIT dye could not confirm that the D-glucose incubated crystals were protein and they were of microscopic size, recreation of those conditions (denoted by a star in Table 10) was not attempted. The most promising condition for $BcsC^{TPR\ 1\text{-}15}$, MCSG-3 G12 (0.1 M sodium acetate pH 4.6, 2 M sodium formate) had spherulite crystal morphology (poor X-ray diffraction quality due to thinness in one dimension over the others), so attempts were made to improve the crystals. First, each component in the initial condition was varied separately (0.05 to 0.2 M for sodium acetate and 0-3 M for sodium formate, respectively) to monitor for improved crystal growth, yet the variations either had no growth or the crystals looked thinner or more brittle. Expansion of the initial crystal condition also involved varying the stock protein concentration as well as crystal drop setup (reservoir buffer to protein stock ratios) and found that initial stock concentrations of 25 to 30 mg/mL and protein solution to reservoir buffer ratios of 3:2 and 2:1 yielded crystals that were larger in all 3 dimensions. The pH of the condition was also varied 2-3 units above and below the initial crystal condition (pH 5.39) but keeping the pH similar to the original condition yielded the best crystal results. To slow down the crystal growth in the one dimension relative to the other two, the incubation temperature of the crystal trays was decreased to 4°C instead of the normal 16°C storage temperature, but this proved to be too dramatic of a change, as no crystal growth was observed at this temperature. Several additives were supplemented into the initial condition (0.1 mM sodium acetate and 2 M sodium formate) such

as dithiothreitol (2 mM DTT), but no crystal formation was seen. The precipitants ethylene glycol and glycerol (2.5, 5, and 10% respectively) were added to the initial condition, which only decreased viability of the crystals due to malformation and fragility. While each of the preceding trials were conducted both with and without streak seeding with a variety of spherulite crystal stocks to introduce microcrystals as nucleation points for continued growth, only after numerous attempts at crystal seeding with a very dilute crystal stock (1/10 to 1/100 dilutions with crystal condition) were rod-like crystals isolated. Example images of the multiple isoforms obtained from these crystal seeding results are displayed in Figure 28, including the refined spherulites (Figure 28A) and the rod-like crystal (Figure 28B). Single crystals isolated from these drops were assessed to have resolutions of ~3.5Å (Figure 29) following X-ray diffraction analysis at the Canadian Light Source (CLS). However, the cryoprotectant (16.5% (v/v) glycerol, 1.32 M sodium formate, 66 mM sodium acetate) used to stabilize these crystals was not ideal and the diffraction pattern was smudged leading to unusable reflections in the diffraction images. Replicating the growth conditions for this crystal in additional trials with a focus on seed quality (*ie.* using the best-looking crystals for successive seeding) and improved cryoprotectants for diffraction (consisting of 3.96 M sodium formate and 66 mM sodium acetate) successfully led to a 3.1Å resolution dataset (data statistics are summarized in Table 12 and an image of the diffraction pattern is depicted in Figure 29). Solvent content analysis was consistent with one copy in the unit cell (2.34 Matthews Coefficient).

Table 12. Statistics for X-ray Data Collection and Processing of BcsC$^{TPR\ 1-15}$

Diffraction Source	CMCF beamline 08B1-1
Wavelength	0.9795
Temperature	100 K
Exposure time	0.2s
Number of images collected	180
Oscillation range	0.2 degrees per image
Space group	C2
Cell dimensions	
***a, b, c,* (Å)**	148.54, 52.62, 91.25
α, β, ɣ, (°)	90, 117.28, 90
Resolution range (Å)	48.88-3.10 (3.21-3.10)
Total number of reflections	38478 (3791)
Total number of unique reflections	11574 (1115)
R_{meas}	0.052 (0.585)
$I/\sigma I$	14.0 (2.60)
R_{pim}	0.028 (0.313)
$CC_{1/2}$	0.999 (0.873)
Completeness (%)	99.6 (99.9)
Redundancy	3.3 (3.4)
Mosaicity	0.26

Attempts to recreate crystal hits with BcsC$^{TPR\ 12-21}$ were successful for the Top 96 A1 (0.2 M MgCl, 0.1 M Tris-HCl pH 8.5, 30% (w/v) PEG 4000) and MCSG-2 F6 (0.2 M sodium malonate pH 7, 20% (w/v) PEG 3350) conditions. Initially, several expansion plates were conducted with the Top 96 condition using different concentrations of buffer (0.1 to 0.3M), and PEG (10 to 30%), which included different ratios of stock protein solution to reservoir buffer in each plate created (similar to that outlined in the last paragraph). The crystals yielded from an

optimized drop (2:1 protein to buffer ratio; Figure 30) were large and promising, yet after diffraction, the resolution was found to be quite poor (12-14 Å). Attempts were made to improve this resolution by harvesting newly formed crystals that were unblemished, yet no improvement was seen in the resolution, which remained between 12-14 Å. Although the exterior appearance of the crystals looked refined, the interior stacking was so poor that it seemed impossible to optimize. Crystals from the MCSG-2 F6 condition were more recently recreated (Figure 31) and X-ray diffraction analysis needs to be performed on this condition to further justify crystal optimization.

A) **B)**

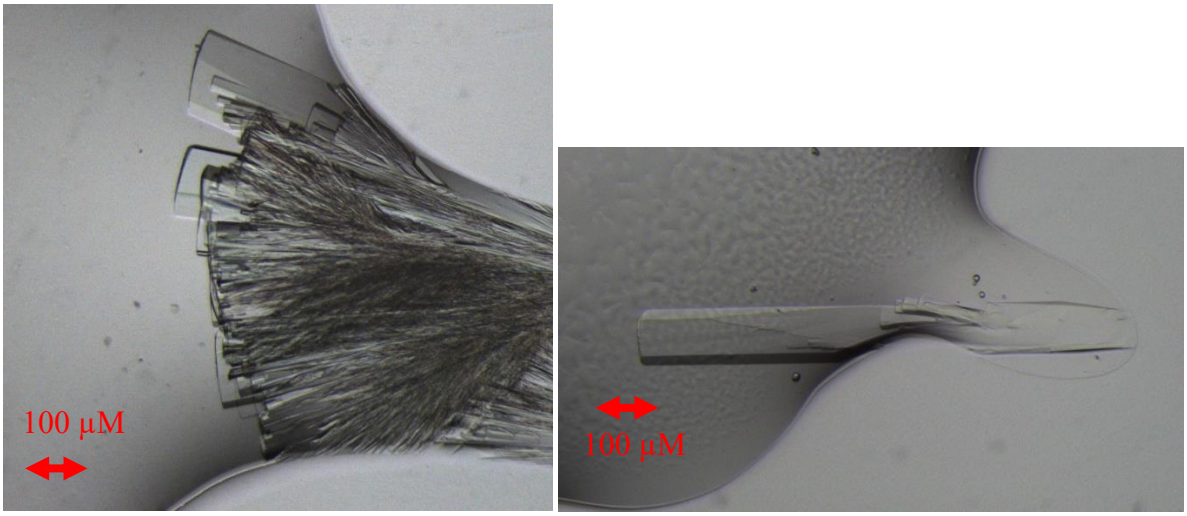


Figure 28. Representive Images of Different Isoforms of BcsC$^{TPR\ 1\text{-}15}$. A) Multiple crystals extending into drop with end available for harvesting with a growth time of approximately 6 months. B) Long rod-like crystal before harvesting with a growth time of approximately 4 months. Each crystal condition contained 0.1 M sodium acetate pH 4.6, 2 M sodium formate.

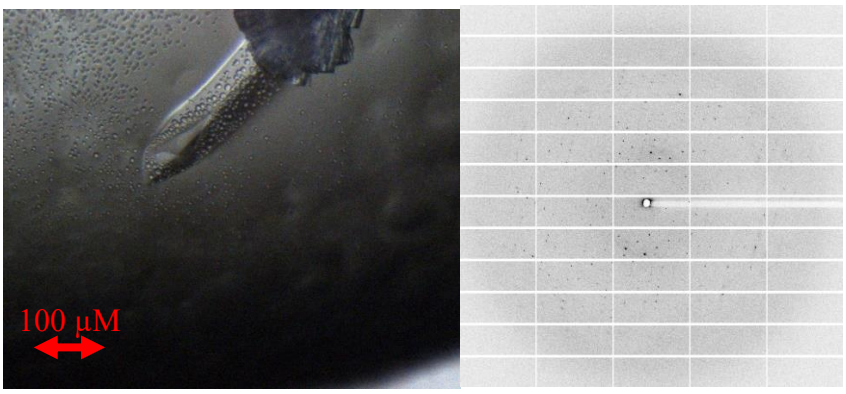


Figure 29. Image of BcsC$^{TPR\ 1-15}$ Crystal Resolution ~3.5Å. Crystal condition contained 0.1 M sodium acetate pH 4.6, 2 M sodium formate. Panel A depicts an image of the crystal growing in the drop prior to harvesting for analysis after a growth time of approximately 4-5 months. Panel B depicts a representative diffraction image.

Figure 30. Representative Crystal Image of BcsC$^{TPR\ 12-21}$. Crystal condition containing 0.2 M MgCl, 0.1 M Tris-HCl pH 8.5, 30% (w/v) PEG 4000 Image of BcsC$^{TPR\ 12-21}$ crystal resolution ~14Å with a growth time of approximately 6 weeks.

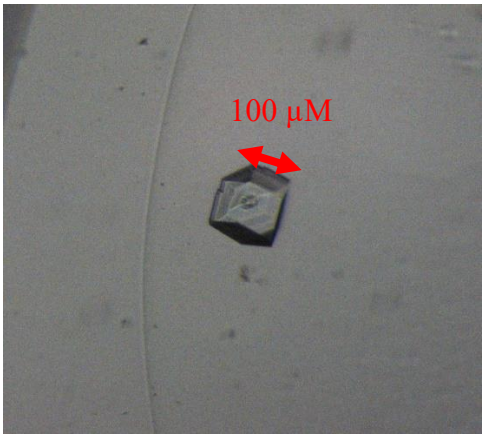


Figure 31. Crystal Image of BcsC$^{TPR\ 12-21}$. Crystal measured at approximately 100 µL and condition containing 0.2 M sodium malonate pH 7, 20% (w/v) PEG 3350 with a growth time of approximately 3 months.

4.5 Molecular Replacement

Molecular replacement was attempted with the BcsC$^{TPR\ 1-15}$ data using a combination of programs across the Phenix and CCP4 online databases. The automated molecular replacement features of these programs (MRage in Phenix and BALBES, MrBump and MoRDa in CCP4 online) performed processes that scan the existing PDB databases for hits based on sequence identity and then perform molecular replacement with the experimental data. Despite multiple trials, the best output with this platform was through the BALBES server (https://ccp4serv7.rc-harwell.ac.uk/ccp4online/) with an R_{work}/R_{free} of 0.5440/0.5470 and a probability of solution of 42.29%. However, attempts to Autobuild with this model were not successful. With the Phenix suite of programs MRage and Phaser were also used in combination with different PDB files from the top five BLASTP and Phyre[2] results (Tables 4 and 5, respectively) along with the PDB files from BcsC$^{TPR\ 1-6}$ (PDB:5xw7) and AlgK (PDB:3e4b). Trimming of each of these models to an alanine backbone was also conducted and attempted as an alternate route to fit the data. The best results so far have consisted of trimming the BcsC$^{TPR\ 1-6}$ model to an alanine backbone with

2 copies in the asymmetric unit. The output statistics have indicated that this solution has a log likelihood gain (LLG) of 48.24 and a translation function z score (TFZ) of 4.5. Rounds of Autobuilding in Phenix using Resolve were attempted with this and other models and the most successful outputs have led to structural models with an R_{work}/R_{free} of 0.3920/0.5827 that are not good candidates to pursue given the poor fit of the model to the density (as evidence in Figure 32).

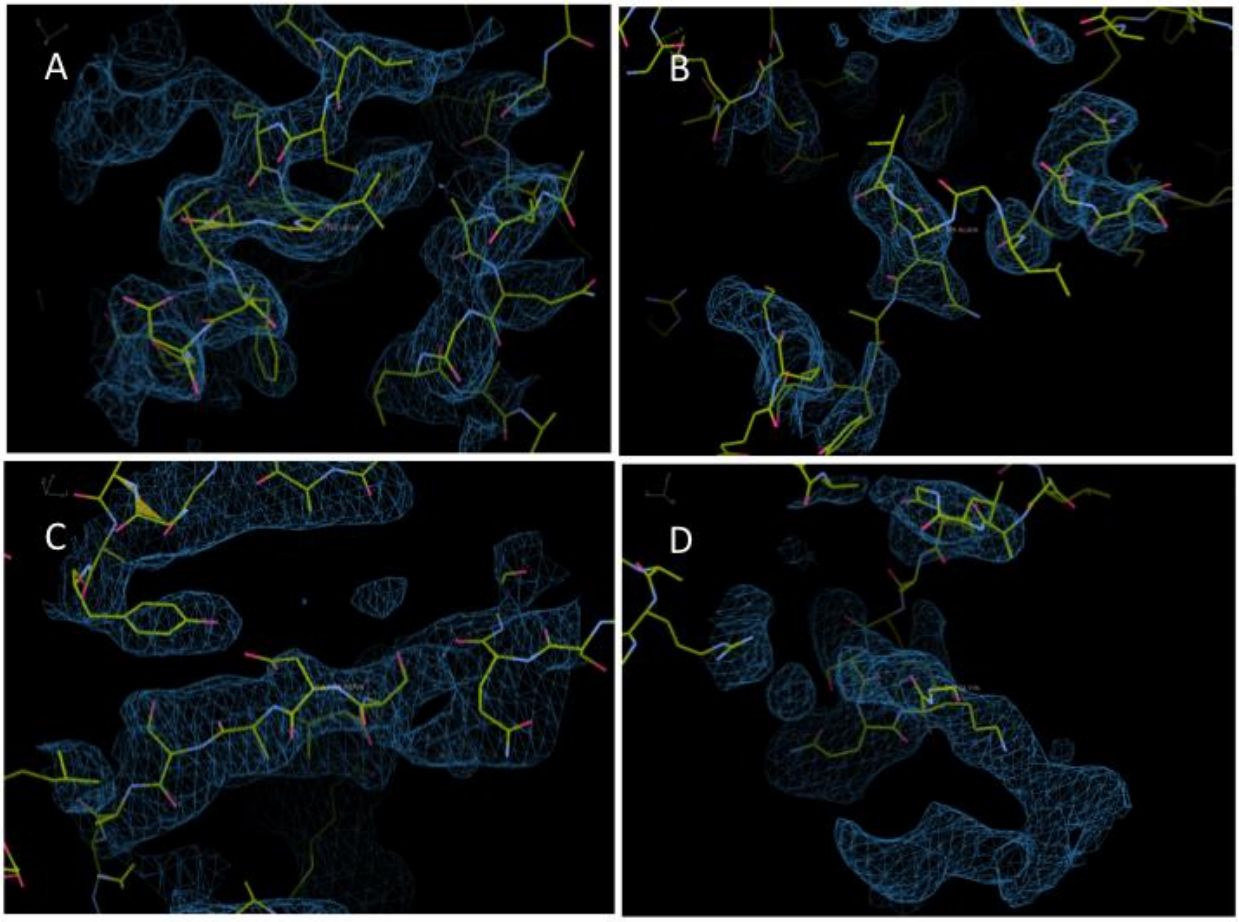


Figure 32. Representative Images of the Molecular Replacement and Autobuild Output of $BcsC^{TPR\ 1\text{-}15}$. Images were rendered in Coot where the blue mesh represents the electron density (rendered at 2 sigma) to which the structural model (yellow backbone) has been fit. Panels A-D depict four separate regions of the single chain modeled into the asymmetric unit.

4.6 Selenomethionine Expression, Purification, and Crystallization

The introduction of selenium into a protein via supplementation of selenomethionine is a common method when attempting to solve the phase problem using heavy atoms in crystallography (141). Figure 33 illustrates the SDS-PAGE image from the expression, IMAC purification and anion exchange chromatography of $BcsC^{TPR\ 1\text{-}15}$ with SeMet minimal media. The yield of protein was less (~2.5 mg/L of culture) when compared with expression in rich media (~3-4 mg/L of culture), but enough protein was purified to proceed to crystal trials. Initially, a test expression of 1 L of SeMet culture was conducted, purified, and placed into crystals trays with seeding using the same condition (0.1 M sodium acetate pH 4.6, 2 M sodium formate) diffracted in the native $BcsC^{TPR\ 1\text{-}15}$ and protein crystals successfully formed. Thereafter, a larger 4 L culture expression was performed, and the resultant yield of protein was approximately 2.5 mg/L of pure recombinant protein following IMAC and ion exchange purification (Figure 33, lane 9 and 10; Figure 34, peak at 120 mL). This protein sample was used to set-up multiple replicates of the successful crystal condition for this construct and the early (2 wks old) representative images of the SeMet crystals (Figure 35) indicate a possibility of diffraction quality crystals given enough time.

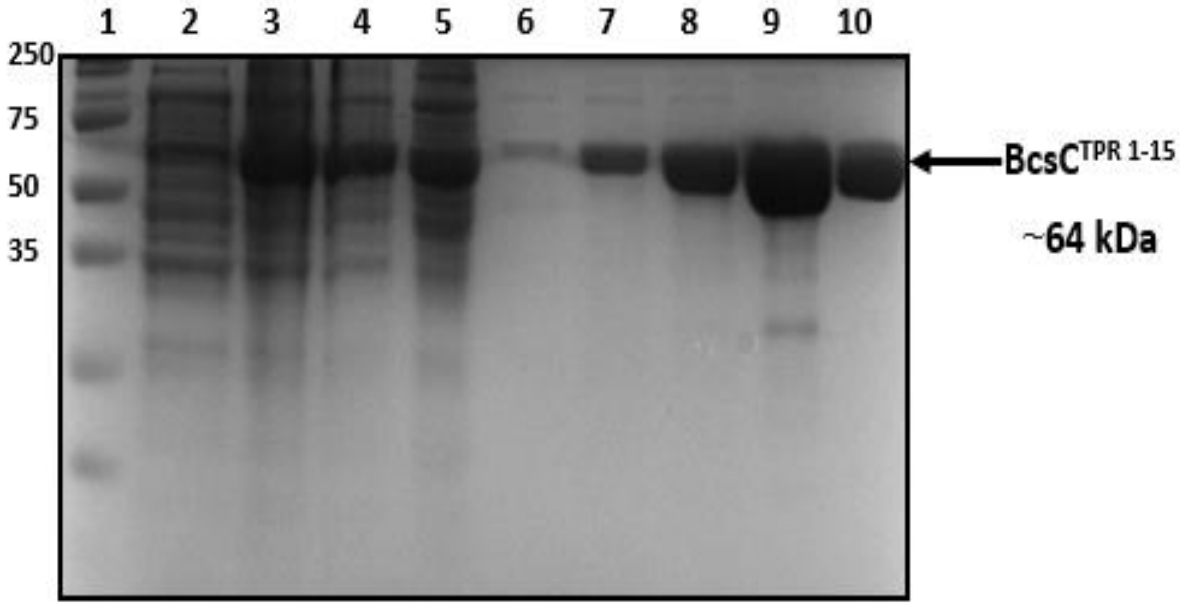


Figure 33. SDS-PAGE Expression and Purification of SeMet BcsC$^{TPR\ 1-15}$. SDS-PAGE analysis (12% (v/v)) of BcsC$^{TPR\ 1-15}$ SeMet minimal media expression of 1 L cultures incubated at 22°C and 160 rpm shaking, Ni-NTA purification fractions, and an anion exchange elution. Lane 1, molecular-weight markers (kDa); lane 2, expression induction time 0 sample (20 mL of sample); lane 3, expression time 4 h sample (15 mL of sample); lane 4 expression time 16 h (10 mL of sample); lane 5, Ni-NTA column unbound lysate; lane 6, wash with lysis buffer (no imidazole); lane 7, wash 1 (20 mM imidazole); lane 8, wash 2 (40 mM imidazole); lane 9, elution buffer (250 mM imidazole); lane 10, fraction from anion exchange chromatography column.

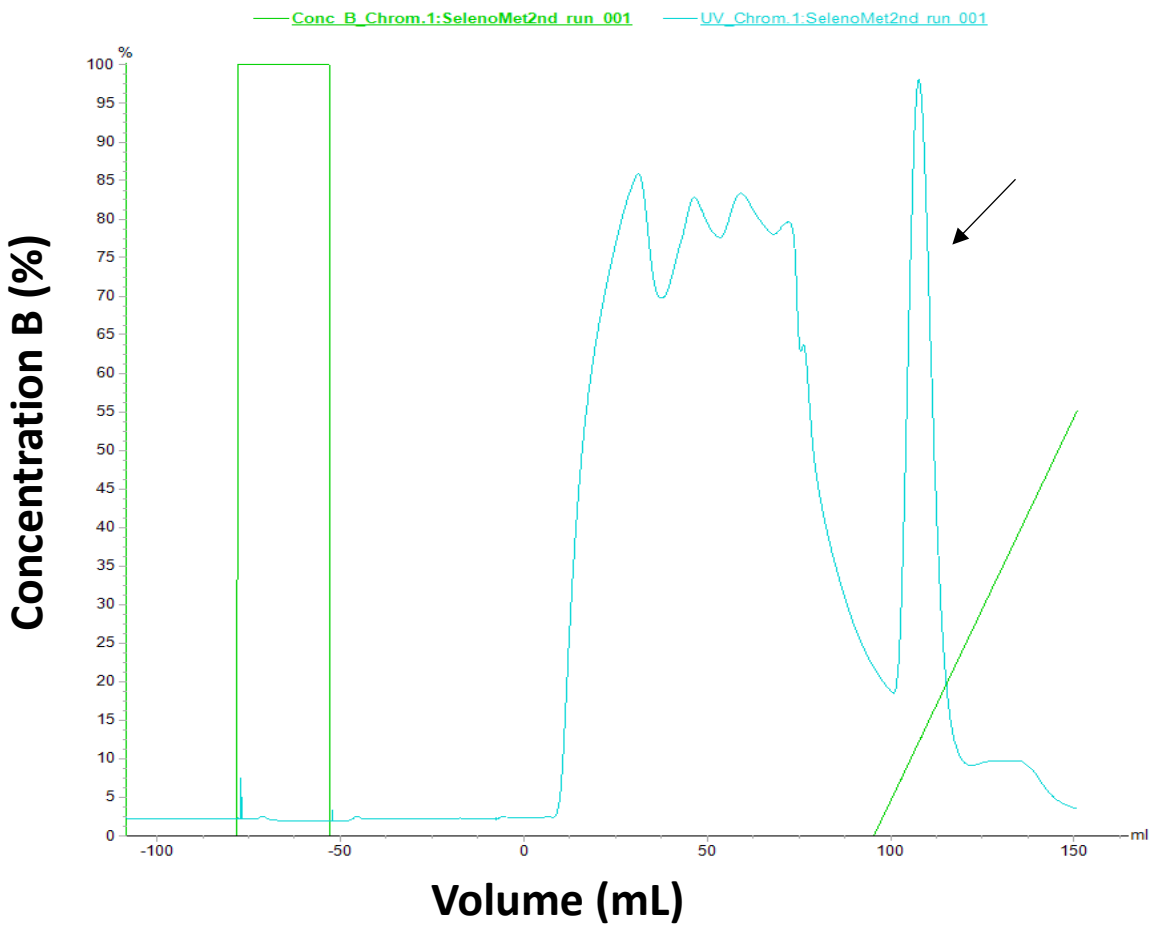


Figure 34. Anion Exchange Purification Profile of SeMet Rich BcsC$^{TPR\ 1\text{-}15}$. The blue line represents mAU for the protein BcsC$^{TPR\ 1\text{-}15}$, where the peak is 355 mAU. The green line represents the percentage of salt eluting. This SeMet containing protein does not produce a typical BcsC chromatograph. The first peak represents the protein passing over the anion column which does not seem to bind to the column until supersaturated.

A) **B)** **C)**

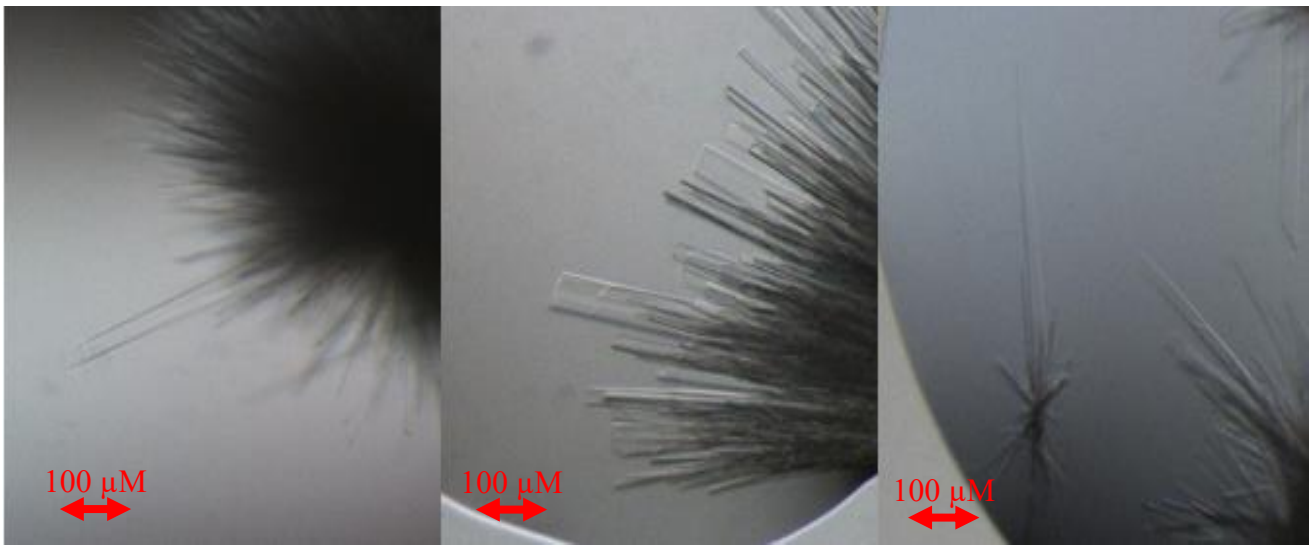


Figure 35. Representive Images of Different Isoforms of SeMet BcsC$^{TPR\ 1\text{-}15}$. Crystal condition containing 0.1 M sodium acetate pH 4.6, 2 M sodium formate with a growth time of approximately 4 weeks.

4.7 Small Angle X-ray Scattering (SAXS)

SAXS analysis was conducted using Frameslice to merge the data, followed by SCATTER to analyze and fit the data patterns. Afterwards, the program PRIMUS was used to fit the P(r) function for later processing with online servers (ATSAS and SIBYLS). The initial results of manual data fitting using SCATTER and later automated online processing using ATSAS is presented in Table 13. The Rg value is the radius of gyration of a particle and can be obtained from Guinier fitting both manually or using an auto Rg function (128). The Rc value of a particle is the cross-sectional radius and when used in conjunction with the Rg value it can yield information about the shape of a particle. For example, an elongated particle such as BcsC would be expected to have an Rg value twice as big as the Rc value because the protein particle is expected to be longer than wider. A globular particle would be expected to have similar Rg and Rc values. These trends of an elongated protein agree with the values collected across all BcsC

constructs, which all have at least double the Rg value compared to Rc. The D_{max} (Å) is the maximum diameter of the particle and in the case of BcsC, the maximum diameter was similar in all four constructs. However, one should exercise caution when interpreting the data as *ab initio* shape reconstruction requires monodisperse data (142), but the chi-squared values of the raw data for $BcsC^{TPR\ 9\text{-}21}$ and $BcsC^{TPR\ 12\text{-}21}$ constructs (1.36 and 0.46, respectively) support that there is decent data for fitting and interpretation. Furthermore, when solving the inverse Fourier transform of the scattering profile, the D_{max} is an adjustable parameter and can be highly sensitive to sample quality and is difficult to predict with accuracy (142, 143). It should also be noted that in cases where a scattering particle has flexibility it may be difficult to choose a D_{max} value, which may apply to BcsC, a protein that has been reported to have a flexible hinge region (90). The lowest D_{max} value (117 Å) was from $BcsC^{TPR\ 9\text{-}21}$ and the model of this construct (as illustrated in Figure 36) adopts a bend or U-shaped fold and has a smaller value to that of the shorter $BcsC^{TPR\ 12\text{-}21}$ construct, which has a fully elongated model (depicted in Figure 37) and the highest D_{max} value (128 Å). It may be that the discrepancy between these values is an indication of greater flexibility within the longer $BcsC^{TPR\ 9\text{-}21}$ construct.

The $BcsC^{TPR\ 1\text{-}11}$ and $BcsC^{TPR\ 1\text{-}15}$ SAXS models that encompass the region of the known $BcsC^{TPR\ 1\text{-}6}$ crystal structure (PDB:5xw7) were manipulated in PyMOL to overlay the structures using DAMAVER. $BcsC^{TPR\ 1\text{-}6}$, in each of the two overlaid models (Figure 38 and 39, respectively) fit the TPR 1-6 regions in a lobe at the bottom of the SAXS model in an orientation where the C-terminus of $BcsC^{TPR\ 1\text{-}6}$ faces towards the unoccupied space of the SAXS model and the N-terminus of the elongated structure is at the other end of the lobe. We deemed this as an important quality control check for the fitting, as the unaccounted-for space in the SAXS model was due to the missing TPR regions (*ie.* TPRs 7 and above) in the model that would be C-

terminal to TPR 6. As expected, BcsC$^{TPR\ 1\text{-}15}$ does appear slightly larger overall in size, but interestingly has a tighter fit of the N-terminus to the BcsC$^{TPR\ 1\text{-}6}$ model. However, these models are predictive in nature and further studies involving other structural techniques would need to be done to correlate these differences in N-terminal flexibility of conformations.

Table 13. SAXS Analysis Values from SCATTER and ATSAS

	Scatter Values		ATSAS Values	
Construct	**R_g Value**	**R_c Value**	**R_g Value**	**D_{max} (Å)**
BcsC$^{TPR\ 1\text{-}11}$	37.6	17.6	38.51	125.5
BcsC$^{TPR\ 1\text{-}15}$	49.1	22.1	37.87	122.5
BcsC$^{TPR\ 9\text{-}21}$	46.9	19.8	36.5	117
BcsC$^{TPR\ 12\text{-}21}$	38.1	17.3	37.86	128

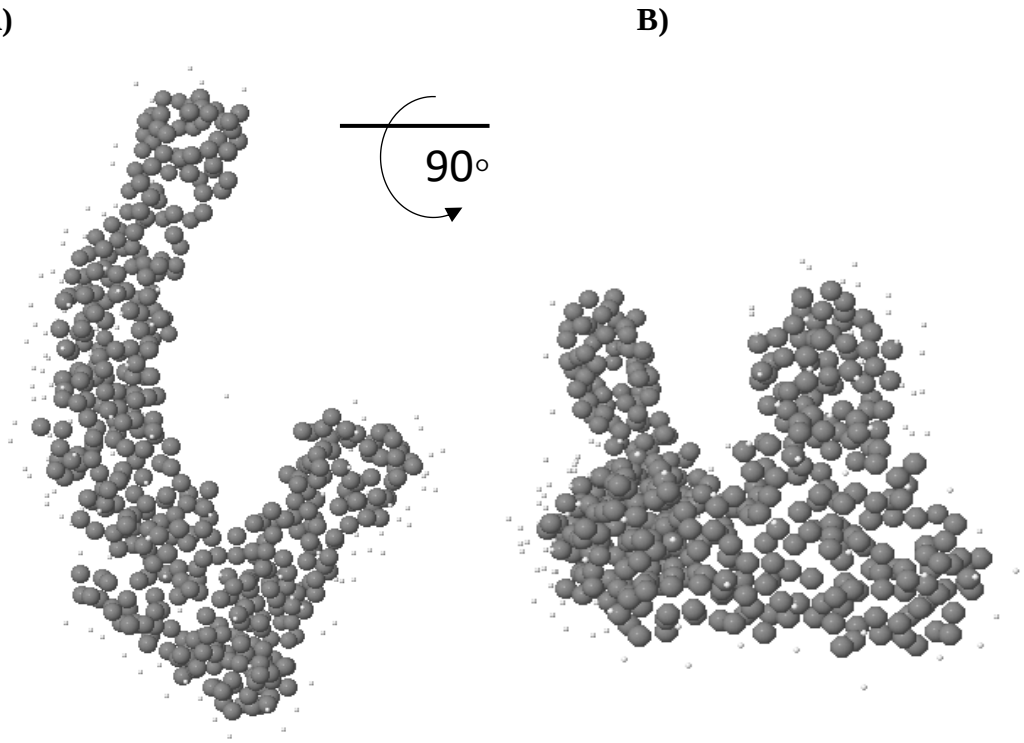


Figure 36. GASBOR Rendering of BcsC$^{TPR\ 9\text{-}21}$. Ab initio spherical representation of BcsC$^{TPR\ 9\text{-}21}$ constructed of dummy residue models displaying a view of each axis (A and B). All models processed on the ATSAS server.

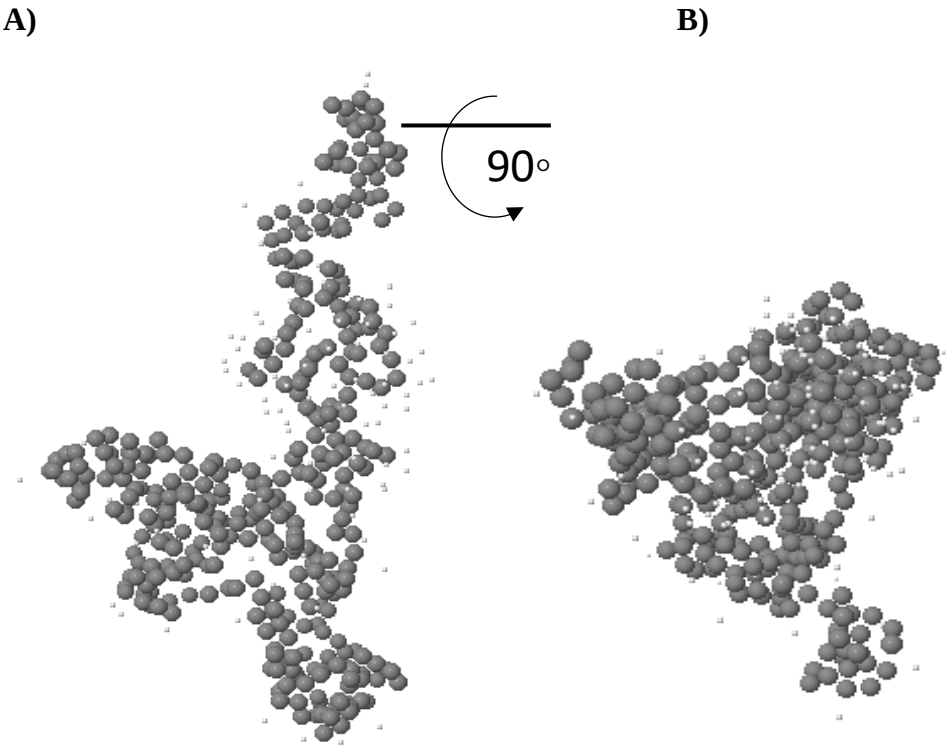


Figure 37. GASBOR Rendering of BcsC$^{TPR\ 12\text{-}21}$. Ab initio spherical representation of BcsC$^{TPR\ 12\text{-}21}$ constructed of dummy residue models displaying a view of each axis (A and B). All models processed on the ATSAS server.

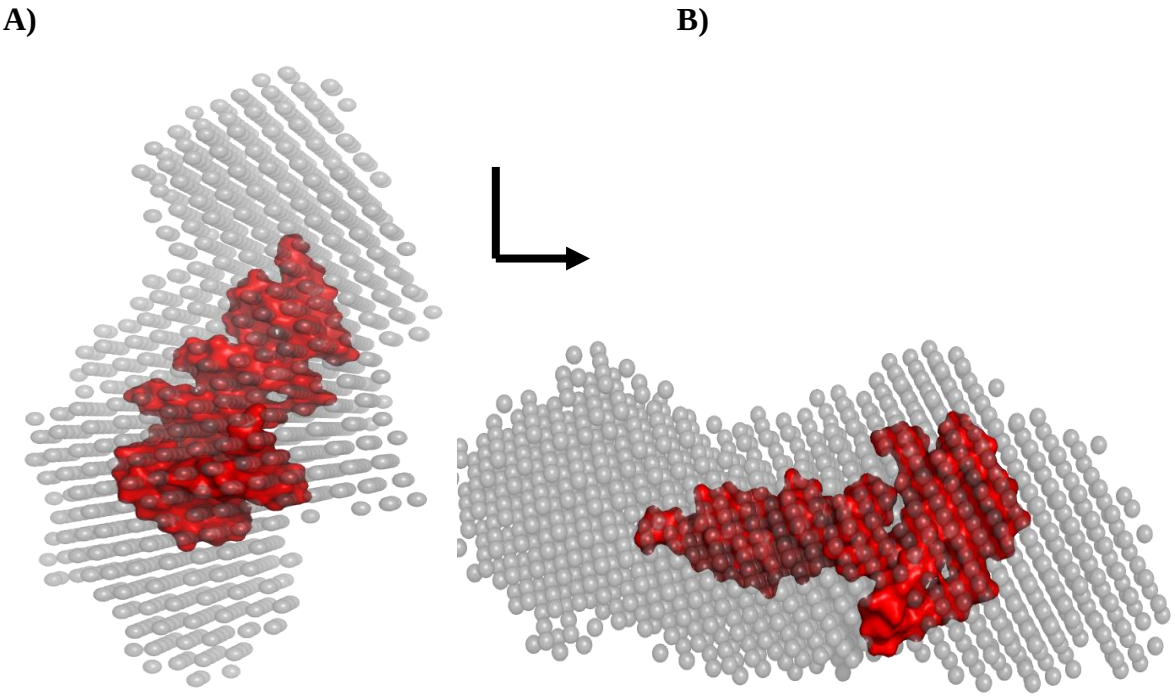


Figure 38. Structural Fitting of BcsC$^{TPR\ 1\text{-}11}$ SAXS Model with BcsC$^{TPR\ 1\text{-}6}$ PDB Model. PyMOL rendered fitting of DAMAVER rendered SAXS model of BcsC$^{TPR\ 1\text{-}11}$ with poly-alanine structure of BcsC$^{TPR\ 1\text{-}6}$.

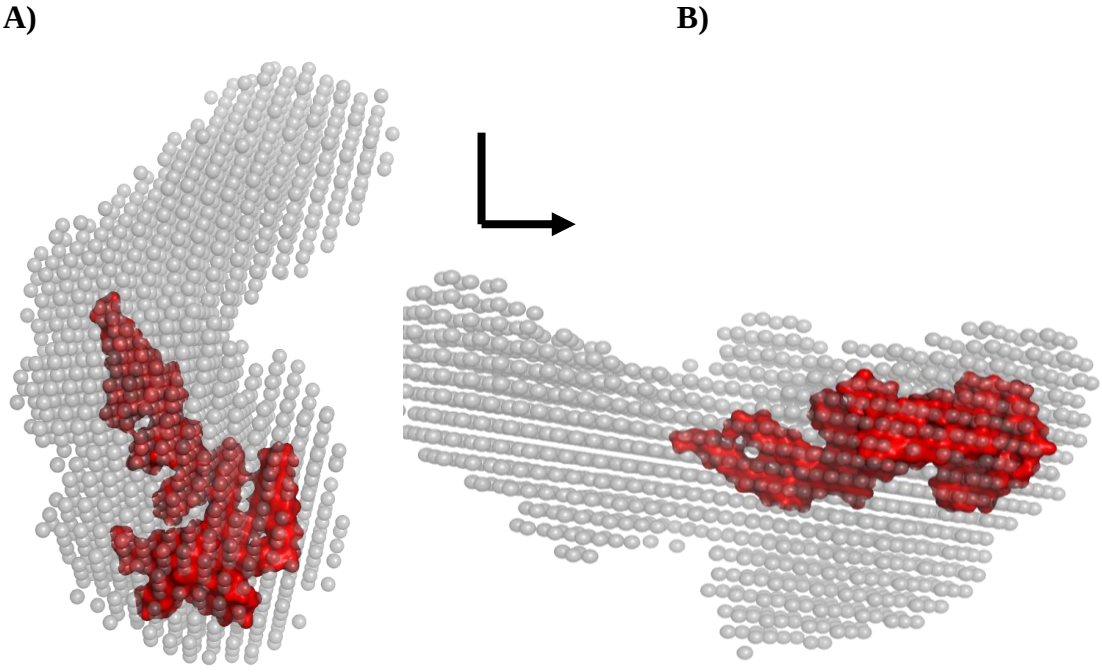


Figure 39. Structural Fitting of BcsC$^{TPR\ 1-15}$ SAXS Model with BcsC$^{TPR\ 1-6}$ PDB Model. PyMOL rendered fitting of DAMAVER rendered SAXS model of BcsC$^{TPR\ 1-15}$ with poly-alanine structure of BcsC$^{TPR\ 1-6}$.

5. Discussion

5.1 Bioinformatics

Bioinformatics provided a useful framework for the planning of downstream experimentation with programs like ProtParam, but also was pivotal for analyses with SAXS, molecular replacement and SeMet-derivative generation. For example, the number and position of methionine residues is pertinent information for SeMet incorporation into each construct as there must be enough to make structure solution by this route plausible. The phasing power of a heavy atom derivative is dependent upon the resolution of the crystal structure and the size of the protein, as a higher resolution and a smaller protein increases the probability of success (141), but a general rule of thumb is to have one SeMet incorporated for every 75-100 amino acids (144), which BcsC narrowly satisfies. One notable characteristic of BcsC highlighted through the program ProtParam was the absence of cysteine residues, which means BcsC is unable to form a disulphide bond between protein chains (145). Disulphide bonds can be important to the natural stability of the protein (i.e. folding) or in the case of oxidative-reductive cycling, the disulfide bond may be important for a proteins activity (145). The Dsb proteins facilitate the correct formation of disulphide bonds for proteins in the periplasm should this be necessary (145). However, the lack of cysteine residues (and disulphide bonds) may suggest that BcsC is meant to be flexible in nature. The formation of cysteine bonds could limit its ability to facilitate interactions with other proteins in the periplasm or even cellulose chains.

A subsequent program that was used frequently in bioinformatics searches, Phyre^2, was also used to create hypothetical structural models of the entire TPR domain, as well as all 6 BcsC

constructs. Figures 9-15 were all based on proteins with some homology at the amino acid level to BcsC. All the structural models identified had TPR helices and/or contained a high number of helices with suggested involvement in protein-protein interactions (146–150). For example, the structure of O-linked N-acetylglucosamine transferase contained 11.5 TPR repeats that formed an elongated superhelix suggested to be a molecular scaffold for other proteins, which in a larger complex facilitates glycosylation of other proteins (148). This functional homology coincides with the assumed BcsC TPR domain interactions with other periplasmic proteins, such as BcsZ and BcsA-BcsB, and also with the closely related alginate system in which the TPR protein AlgK is suggested to have interactions with the other periplasmic proteins AlgX, Alg44, and AlgE (74, 151, 152). Although these modelling results are interesting, the proteins that had higher amino acid identity returned through BLAST searches (Table 4) were also deemed relevant, as Phyre^2 can sometimes miss newly deposited structures in the PDB database. Given that amino acid identity is a good indicator for the success of molecular replacement models, combining the PSI-BLAST iteration with Phyre^2 was a good strategy, which was designed to increase the probability of obtaining a successful molecular replacement for structure solution. All BcsC constructs also contained a percent identity of between 22 and 25% with the amino acid sequence of AlgK, and between 71 and 73% with BcsC$^{TPR\ 1\text{-}6}$ (as assessed through Clustal Omega alignments in our bioinformatics analysis pipeline), so they were included as molecular replacement modelling options as well.

Disorder prediction software was used to search for disordered regions amongst BcsC protein constructs, as proteins that contain disordered regions can be difficult to purify and crystalize (153–155). Unstructured or disordered regions can direct functions within a protein (156) but can interfere with ordered packing for crystallization. Therefore, disorder prediction

can prove pivotal when creating protein constructs to ensure proper study of both ordered and disordered regions (157). Because each disorder prediction algorithm has its weaknesses (109), a meta approach was taken to study disordered regions. Both metaPrDos and DisMeta contain several different programs (158, 159) and the combined results for all BcsC constructs indicated that there may be considerable disorder at each terminus, which could be explained in part by the His_6-tag. However, as two of the six protein constructs were crystalized, it appears the disorder at the termini had a minimal effect, if any, on protein folding. In addition, $BcsC^{TPR\ 4\text{-}21}$ contains a short series of disordered amino acids predicted slightly above the 5% false positive confidence scale in the middle of the protein. While this region may be insignificant for a protein as large as $BcsC^{TPR\ 4\text{-}21}$ (74.1 kDa), since soluble protein was consistently produced, it has yet to be determined if it played a role in crystal packing as the structure of this construct remains unsolved.

5.2 Protein Expression and Purification

5.2.1 Protein Expression

The first objective of this thesis was to express and purify each of the protein constructs in our panel of BcsC derivatives. A major part of this objective was spent on optimizing the expression conditions for each of the BcsC protein constructs by surveying IPTG concentrations, growth temperatures and the length of expression. A general consensus across the protein constructs $BcsC^{TPR\ 1\text{-}8}$, $BcsC^{TPR\ 1\text{-}11}$, $BcsC^{TPR\ 1\text{-}15}$, $BcsC^{TPR\ 4\text{-}21}$, $BcsC^{TPR\ 9\text{-}21}$, and $BcsC^{TPR\ 12\text{-}21}$ indicated that optimal conditions involved inoculation at 37°C, induction with 1 mM IPTG and incubation at 22°C for 16 h while shaking at 200 rpm throughout the culture period. During preliminary research expression was conducted at 37°C, yet it was later determined that lowering the temperature increased the yield of soluble protein. This was consistent with published

literature that found lowering the temperature and IPTG concentration may slow the translation machinery and give expressed protein more time to fold properly; thereby, increasing soluble protein yields (160). While some constructs fared slightly better with moderate changes, the general expression protocol was important for ease of use when multiple constructs were being expressed and purified simultaneously or in close succession (which was frequently). Thus, with respect to the first thesis objective, all protein constructs were successfully expressed.

5.2.2 Protein Purification

Following expression, a second part of objective 1 was to subject each of the protein constructs to a platform of purification steps to yield sufficient quantities of soluble protein for objectives 2 and 3. The first protein purification step conducted for all recombinant protein constructs was IMAC with Ni-NTA resin beads as each vector was engineered with a His_6-tag to facilitate this type of purification. All affinity tags have the potential to interfere with biological activity or crystallization of a protein, the advantages of a His_6-tag are that it is small, relatively cheap, can be regenerated many times, contains a large binding capacity, operates well under denaturing conditions, and can interact with multiple chromatographic matrices (161). This may have contributed to the successful completion of protein purification through the IMAC stage for all constructs using a general protocol that involved using increasing amounts of imidazole to wash the contaminants and finally elute the target recombinant protein. For all constructs a further purification step was needed, as SDS-PAGE analysis of BcsC indicated numerous protein contamination bands at multiple weights across all constructs. For secondary purification, ion exchange chromatography was conducted and yielded a purity of over ~95% (as judged by SDS-PAGE analysis of the fraction).

While this research was able to successfully express and purify each of the 6 BcsC derivatives, the constructs did not purify to the same degree and amount. BcsC$^{TPR\ 1\text{-}11}$ and BcsC$^{TPR\ 12\text{-}21}$ consistently yielded the largest amounts of purified protein (~6-7 mg/L culture; Table 7) with our standard two-step purification process (eg. IMAC followed by ion exchange). BcsC$^{TPR\ 4\text{-}21}$ (~2.5-3 mg/L culture), BcsC$^{TPR\ 9\text{-}21}$ (~2.5-3 mg/L culture), and BcsC$^{TPR\ 1\text{-}15}$ (~3-4 mg/L culture) all had lower average yields, but still ample amounts for downstream analyses. Of these five constructs, there may be many reasons for different yields between them. One possibility is the construct size, since the two greatest yields came from constructs BcsC$^{TPR\ 1\text{-}11}$ and BcsC$^{TPR\ 12\text{-}21}$ that expressed proteins of the lowest molecular mass (~47.3 kDa and ~40 kDa respectively); whereas the lowest yields came from BcsC$^{TPR\ 4\text{-}2*1}$ (~74.1 kDa), BcsC$^{TPR\ 9\text{-}21}$ (~52.4 kDa), and BcsC$^{TPR\ 1\text{-}15}$ (~64 kDa). Other researchers have noted that the chances of successfully expressing soluble proteins at molecular weights over ~60 kDa decreases significantly due to the propensity to precipitate into inclusion bodies and/or misfold (140). However, in our case, by combining several expression pellets, the amount of protein harvested was sufficient for all downstream applications. Contrary to the size argument, BcsC$^{TPR\ 1\text{-}8}$ was successfully expressed and purified, but was not pursued further since multiple rounds of optimization did not lead to protein with a high degree of purity and stability. However, the complement of five constructs that were used further encompassed the 1-8 TPR region, so we felt that this region would be adequately covered in the repertoire of the 5 constructs we were successful with.

5.3 Dynamic Light Scattering

DLS was successfully utilized to monitor the protein polydispersity in efforts to improve the suitability of buffers for downstream applications, mainly crystallization. Early results for protein constructs BcsC$^{TPR\ 1\text{-}15}$ and BcsC$^{TPR\ 12\text{-}21}$ yielded very aggregated and polydisperse

samples (Figure 25) that also did not crystalize well. However, the polydispersity character of these samples was overcome by varying several important factors. For example, results indicated that polydispersity levels were significantly affected by increasing the purity of the protein sample, conducting DLS experiments in a time sensitive manner, and varying the sample concentration of the protein used in the DLS experiments. The most successful DLS trials were conducted with protein newly purified from an anion exchange experiment that was filtered and adjusted to an approximate concentration of 1-2 mg/mL (Figure 26A & 26B). Following the adoption of some these conditions, there was a direct improvement in the crystallization propensity of the $BcsC^{TPR\ 1\text{-}15}$ and $BcsC^{TPR\ 12\text{-}21}$ protein constructs.

5.4 Protein Crystallization

As part of the second objective of the present research, the stability and composition of each of the purified proteins was optimized for crystallization trials. This step was necessary as protein crystallization is most successful at high levels of purity and conducted under as many different conditions as possible (154), even though this does not guarantee protein crystals will form. To maximize the chances of success, the addition and optimization of the salt concentration is important since salts can aid macromolecules in associating with one another, through competition with proteins for water molecules to fulfill their electrostatics requirements (154). Across several of the recorded crystal hits for BcsCs conditions (Table 10 & 11) sodium can be found. This may be due to a combination of favourable protein-ion interactions or the dehydration effect previously mentioned. From the literature, $BcsC^{TPR\ 1\text{-}6}$ was crystallized to a resolution of 3.27 Å in a condition that contained 100 mM MES (pH 6.2), 3.5 M sodium chloride (90). When combining these results, it seems clear that BcsC constructs behave favourably when

interacting with sodium and future attempts at crystallization should examine sodium-containing conditions more closely.

BcsC was successfully crystallized in two TPR overlapping constructs ($BcsC^{TPR\ 1\text{-}15}$ and $BcsC^{TPR\ 12\text{-}21}$) in several conditions. Although one of the concentrations was only 13 mg/mL, the majority of BcsC crystals hits came at concentrations of 29 mg/ml and above, which was consistent with the aforementioned $BcsC^{TPR\ 1\text{-}6}$ that diffracted at 60 mg/mL (3.27 Å). Of the four crystal hits for $BcsC^{TPR\ 12\text{-}21}$, only the two most promising conditions were reproduced. Multiple trials were conducted to reproduce and optimize a rod-like crystal from the first condition (0.2 M MgCl, 0.1 M Tris-HCl pH 8.5, 30% (w/v) PEG 4000) using grid style screening with variations of buffer, PEG, salt and protein concentrations. While over a dozen promising crystals (large and good three-dimensional character) were harvested and sent for diffraction, the diffraction quality from this crystal form was consistently poor and ranged between 12-14 Å (Figure 30). Multiple attempts were made to harvest fresher crystals for diffraction, but the results were unchanged, and the condition was abandoned in favour of other constructs.

The second promising crystal condition (0.2 M sodium malonate pH 7, 20% (w/v) PEG 3350 (MCSG-2 F6) with $BcsC^{TPR\ 12\text{-}21}$ took far longer to replicate. This crystal required months to grow and only recently has a small cubic crystal (Figure 31) been identified in the condition. This crystal does appear promising but future experimentation to assess the diffraction quality of the crystal (and potential further optimization) still needs to be performed, which is beyond the scope of this research due to time constraints.

Our moderate success with $BcsC^{TPR\ 12\text{-}21}$ crystallization has proved time consuming and encountered multiple hurdles, but this is not unique for the crystallization of TPR containing proteins. Researchers of the homologous TPR containing protein AlgK were met with difficulties

crystallizing and switched to a different organism for crystallization (74). Also, the group involved in crystallizing $BcsC^{TPR\ 1\text{-}6}$ also attempted to solve the structure of $BcsC^{TPR\ 1\text{-}17}$ but were unable to attain a high enough diffraction quality (7–8 Å) (90). From these results it is clear that the characterization of TPR export proteins is extremely difficult, which may be due to the flexible nature of TPRs in general (87, 90) and that continued optimization of $BcsC^{TPR\ 12\text{-}21}$ and other constructs is still worthwhile.

For $BcsC^{TPR\ 1\text{-}15}$, three crystal conditions were identified in initial trials (96-well, hanging drop, Gryphon-setup plates) as having microcrystals worth pursuing. However, for two of the crystal conditions (10% (w/v) PEG 4000, 20% (v/v) glycerol, 0.2 M mixture of carboxylic acids and 0.1 M MES/Imidazole and 0.1 M Hepes: NaOH pH 7.5, 20% (w/v) PEG 8000), manual setup in larger chambered plates (24-well plates) utilizing sitting drop kinetics under the same conditions did not lead to usable crystal forms. Instead, the crystals were often far smaller and nucleated growth in a spherulite type of pattern, rather than individual microcrystals. This variance may be due to the difference in sitting versus hanging drop and increased drop volume (nl to μl, respectively) that affected vapor diffusion. In the presence of a glucose additive (to mimic cellulose ligands for the protein), three successful replicates did lead to two micro-rod and small spherulite crystals (Table 10), but upon testing with IZIT dye for protein composition it was still unclear following addition of the dye, due to the small size of these crystal forms. The third $BcsC^{TPR\ 1\text{-}15}$ condition (0.1 M sodium acetate pH 4.6, 2 M sodium formate) behaved similarly to the other conditions at the start with one spherulite crystal in the initial trials. Improvement of the quality of the crystals was not achieved through altering protein concentrations, pH, temperature and additives, but instead through iterations of microseeding of new crystal conditions by the streak seeding method that led to thicker growth of rod-like

crystals (see different isoforms Figure 28) with diffraction quality to 4 Å. Successive seeding attempts over a 4 month period led to even better growth with much larger and thicker rod-shaped crystals with 3.5 Å resolution diffraction (Figure 29). Trials to improve the cryoprotectants with these crystals (glycerol and ethylene glycol switched for high concentrations of sodium formate) led to improved diffraction images (low anisotropy of reflections) and an enhanced resolution of approximately 3 Å. These results may be due to the increase in sodium formate acting as a dehydrant, as well as a cryoprotectant, similar to what has been identified in other studies (154).

5.5 Molecular Replacement

Molecular replacement was conducted using multiple programs through the Phenix and CCP4 online databases as automated pipelines and with targeted PDB files from Phyre[2] and BLAST searches (Tables 2 and 3 respectively) with favourable homology to BcsC. In addition, the PDB files from $BcsC^{TPR\ 1\text{-}6}$ and AlgK were also used in various forms and combinations with each other as well as the other PDBs. The best solution obtained was with an alanine trimmed variant of the $BcsC^{TPR\ 1\text{-}6}$ structure that had LLG and TFZ molecular replacement statistics of 48.24 and 4.5, respectively. Despite these lower statistical ratings, rounds of Autobuilding in Phenix were attempted with this and other models and the most successful outputs have led to structural models with an R_{work}/R_{free} of 0.3920/0.5827, which is indicative of possible over-refinement of the model given the divergence of the two parameters. Ideally, both the R_{work} and R_{free} should decrease in value together since the R_{free} represents the validation set of data that has not been manipulated and is an indication of whether the model fits the original dataset with any statistical confidence. While future work with molecular replacement will continue, as new models and structures are solved or made available constantly, the complimentary alternative to

this method, SeMet-derivative crystals, was also attempted. Future work could also consist of heavy metal soaking into the crystals to use these atoms as a phasing method to solve the structure of BcsC.

Even in the absence of a resolved structure, the diffraction data for $BcsC^{TPR\ 1-15}$ have supported previously reported data. For example, the unit cell values of the crystal data that has been collected suggest an extended structure in one dimension (148 Å) that is similar to other TPRs. The previously reported crystal structure and SAXS data of $BcsC^{TPR\ 1-17}$ also found that this construct covered 120 Å in length (90). These researchers further suggested (90) that this may be enough to span the periplasm, which is proposed to be 150-180 Å (163) in length. Thus, future work with our crystal constructs will shed more light on this theory.

5.6 Selenomethionine Crystallization

SeMet incorporation into a protein is a common method for solving the phase problem in crystallography (141) and was conducted in this study. Expression yields with the SeMet derivative of $BcsC^{TPR\ 1-15}$ led to slightly lower than native yields, which was likely due to the reduced nutrient load in the minimal media, but still resulted in enough protein for crystal trials that had a higher level of purity (see Figure 19 and 32 for native and SeMet derivatives, respectively). SeMet derived protein was successfully crystallized using seeding with $BcsC^{TPR\ 1-15}$ in the exact same condition as the native $BcsC^{TPR\ 1-15}$ (0.1 M sodium acetate pH 4.6, 2 M sodium formate with a final buffer pH of 5.4). Comparison of the crystal images of the native (Figure 29) and the SeMet protein derivative (Figure 35) crystals clearly show that the native crystals are thicker. However, this difference likely has to do with time since the native crystals have grown for twice as long (2 months as opposed to 1 month for the SeMet derivative). The published structure of $BcsC^{TPR\ 1-6}$ followed this seemingly same trend as the SeMet variant

diffracted to only 3.4 Å, while the native protein was 3.27 Å (90). In a separate study, the homologous export protein AlgK was solved using SeMet phasing (2.8 Å), where the native form was also crystallized to a higher resolution (2.5 Å). The success of both groups of researchers is promising, although more time may be required to develop SeMet variant crystals, so monitoring of the plates is ongoing and will be followed by testing the X-ray diffraction quality of the crystals once they are larger.

5.7 Small Angle X-ray Scattering (SAXS)

SAXS is a technique that can apply to a wide range of particle shapes and sizes that has quickly become a major tool for characterizing macromolecular systems (128) and was used during the present study. The amount of diffraction data generated by a SAXS analysis is cumbersome to manipulate with the mathematical tools and algorithms available but can produce data that provides some valuable model interpretative results. When used in conjunction with other techniques, such as X-ray crystallography, the results of SAXS can be even more powerful by correlating evidence between two structural models. The current project was not able to provide a structure to balance the results from SAXS, so a structure of $BcsC^{TPR\ 1-6}$ from the PDB was used instead. Initial SAXS data (Table 13) was successful in generating models for the 4 tested BcsC constructs as elongated particles (as opposed to globular), which was consistent with what is known of the composition of BcsC. When the P(r) function was plotted, a long protein consisting of dummy residues was constructed for all 4 constructs, of which only the two C-terminal constructs were presented ($BcsC^{TPR\ 9-21}$; Figures 36 and $BcsC^{TPR\ 12-21}$; Figure 37), as the two N-terminal constructs ($BcsC^{TPR\ 1-11}$; Figures 38 and $BcsC^{TPR\ 1-15}$; Figure 39) were displayed with the PDB of $BcsC^{TPR\ 1-6}$ in a fitted model instead. When visually comparing the two models it is clear that $BcsC^{TPR\ 1-15}$ is slightly bigger than $BcsC^{TPR\ 1-11}$ and seems to have a closer fit. In a

recent study, SAXS data from $BcsC^{TPR\ 1-17}$ was presented and suggests that $BcsC^{TPR\ 1-17}$ is 120 Å in length (90). Since the bacterial cell envelope is believed to span 150-180 Å (163), this allowed for the theory that full length of the TPR region of BcsC may span the entire periplasm; thereby, connecting BcsB to the ß-barrel domain of BcsC to facilitate the export of cellulose (162). Our current research also supports this theory given that the conservative estimates of D_{max} values (approximately 120 Å) and the unit cell values from the diffraction data collected ($BcsC^{TPR\ 1-15}$ = 148 Å) suggest there is enough length in these models to cover even more of the periplasmic space than previously suggested in the literature (163).

5.8 Summary and Significance

Bacteria can adapt using many strategies for survival across a range of environments and one of the most prevalent strategies in bacterial infections is to utilize biofilms, thereby, making research on exopolysaccharide secretion systems vital for the prevention of and controlling the spread of biofilms. Without the presence of exopolysaccharides in a biofilm, the antibiotic resistance decreases dramatically (68, 164, 165). Although the study of bacterial cellulose secretion systems has been important in industrial fields, such as cosmetics, drug delivery, and food processing (166), it is also crucial in the medical field, due to its role in pathogenesis with respect to *Escherichia.coli* (22), *Salmonella* spp. (4), and *Pseudomonas aeruginosa* (23). The present research was conducted to help bridge the understanding between current exopolysaccharide secretion systems present in enteric bacteria and areas of the cellulose biosynthetic system that are less defined. Specifically, this research focused on elucidation of key details about the bacterial biosynthesis protein BcsC.

BcsC is predicted to be responsible for export of cellulose from the cytoplasm to the extracellular space and presents a significant target for inhibition of cellulose production by the

cell; which, in turn, could aid in prevention and control of biofilms (47). BcsC has functional (162), structural (74, 90), and sequence (22 to 25% identity) homology to AlgK, a previously characterized protein involved in exopolysaccharide export (74) in the characterized alginate system (162). However, key differences exist between them, for example, BcsC contains several more TPR regions, which may add enough length to span the periplasm (as evidenced by our D_{max} SAXS data) and facilitate direct protein-protein interactions with the synthesis complex comprising BcsA-BcsB. This setup would be novel compared to previously characterized alginate and PNAG secretion systems, but the relevance and mechanism of interactions are still unknown. Currently, a structure of the first 6 TPR regions of BcsC exists that presented multiple conformations and suggests flexibility within the structure (90), yet much is left unanswered due to the resolution of the model, discrepancies in the overall number of predicted TPR regions, and the absence of a complete model that predicts specific functions with respect to the proposed protein-protein interactions. We hypothesized that a structural investigation of the N-terminus of BcsC would confirm that it contains a series of TPR folds and that this region is important to ligand binding. The objectives entailed first expressing and purifying all protein constructs and secondly, performing extensive crystal screening, followed by the refinement of promising conditions that may lead to improved quality for X-ray diffraction data. The final objective was to build a structural model of the BcsC protein constructs through a combination of X-ray crystallography and SAXS experiments.

With respect to objective 1, BcsC was successfully expressed and purified in all constructs, five of which reached a high level of purity (95%) after secondary purification. While $BcsC^{TPR1-15}$, $BcsC^{TPR\ 12-21}$, $BcsC^{TPR\ 1-11}$, and $BcsC^{TPR\ 9-21}$ could be readily purified with our general protocol, $BcsC^{TPR\ 9-21}$, $BcsC^{TPR\ 4-21}$, and $BcsC^{TPR\ 1-8}$ were difficult to work with and

initially abandoned in preference of the others. However, late in the project, work with BcsC$^{TPR\ 9\text{-}21}$ and BcsC$^{TPR\ 4\text{-}21}$ resumed successfully so future research can also focus on downstream analyses with these C-terminal protein constructs. Together, this panel of purified constructs (BcsC$^{TPR\ 1\text{-}15}$, BcsC$^{TPR\ 12\text{-}21}$, BcsC$^{TPR\ 1\text{-}11}$, BcsC$^{TPR\ 9\text{-}21}$ and BcsC$^{TPR\ 4\text{-}21}$) are noteworthy because they span the entire BcsC region and improved our chances for successful crystal hits (due to size or disorder), as was evidenced by the BcsC$^{TPR\ 4\text{-}21}$ and BcsC$^{TPR\ 1\text{-}8}$ constructs. This work is important because it represents the successful report of regions beyond TPRs 1-6 (90) being expressed and purified to near homogeneity. The entire panel of BcsC constructs now provides a unique opportunity to crystallize sub-regions of TPRs or perform functional work that may prove to have specific roles in interacting with BcsA/B and newly formed cellulose (*i.e.* the N-terminal containing TPR constructs) or the outer membrane beta-barrel (*i.e.* the C-terminal containing TPR constructs) as has been inferred in the literature (162). Given that a standard protocol has now been generated to produce protein from 5 of the 6 constructs attempted, the only remaining future directions for this objective would be to improve purification from the last construct by minimizing protein degradation through inclusion of protease inhibitors and the addition of glycerol to purification buffers to begin with.

Our work towards the second objective of optimizing conditions for the crystallization of BcsC constructs was also successful. The construct screening order for crystal trials in the project followed a timeline of BcsC$^{TPR\ 1\text{-}15}$, BcsC$^{TPR\ 12\text{-}21}$, BcsC$^{TPR\ 1\text{-}11}$, BcsC$^{TPR\ 9\text{-}21}$, and then BcsC$^{TPR\ 4\text{-}21}$. This was based on when our expression/purification trials yielded pure protein to work with and led to some crystals undergoing more optimization than others. During crystal screening with each of these constructs, numerous crystals were observed, expanded upon, and sent for diffraction analysis. BcsC$^{TPR\ 1\text{-}15}$ and BcsC$^{TPR\ 12\text{-}21}$, two constructs that together overlap

the entire length of the BcsC TPR domain, were successfully crystallized. The labour-intensive crystal optimization in this thesis involved tens of thousands of different individual trials, in addition to the use of variations in conditions, including wide-ranging protein concentrations and ratios to buffer, as well as hundreds of well replications with crystal seeding. Current successes with crystallization of two overlapping TPR regions provide a framework to build a model that significantly advances the current one for BcsC (90) and other homologous protein models that have not presented the benefit of separate sources of structural overlaps. To achieve this goal, the screening of C-terminal constructs ($BcsC^{TPR\ 4\text{-}21}$, $BcsC^{TPR\ 9\text{-}21}$, and $BcsC^{TPR\ 12\text{-}21}$) should be prioritized along with the crystal optimization of $BcsC^{TPR\ 12\text{-}21}$ (0.2 M sodium malonate pH 7, 20% (w/v) PEG 3350), as an N-terminal construct in the final experimental stages. A common strategy for conducting these experiments would be to first screen any constructs that have not had extensive screening/optimization ($BcsC^{TPR\ 4\text{-}21}$and $BcsC^{TPR\ 9\text{-}21}$), followed by a rescreening without the His_6-tags of $BcsC^{TPR\ 12\text{-}21}$, $BcsC^{TPR\ 4\text{-}21}$, and $BcsC^{TPR\ 9\text{-}21}$ if no promising hits were found in the primary screening (167).

The final objective of providing a structural model of BcsC had a two-pronged approach. The primary method of crystallization has yielded crystals from native $BcsC^{TPR1\text{-}15}$ condition (0.1 M sodium acetate pH 4.6, 2 M sodium formate), that were diffracted to approximately 3 Å. At this resolution, less errors may be inherent compared to the current structural model of $BcsC^{TPR\ 1\text{-}6}$, which is 3.27 Å (90). Multiple molecular replacement models were attempted with different molecular replacement programs, however, despite the fact that the amino acid homology appeared to be high enough to produce a solution (*ie.* greater than 30% in some cases), these efforts were ineffective. As a separate avenue to determine the phases, SeMet derivatives of this protein were expressed and purified. This SeMet preparation also led to crystals when subjected

to the same conditions as the native crystals, but the SeMet crystals are still growing and, due to the time constraints of the thesis, will be pursued as a future direction of this research. However, this avenue to structure solution is promising given that there are up to 8 methionine positions that can be replaced in $BcsC^{TPR\ 1\text{-}15}$, leading to an increased probability that there is a substructure of these heavy atoms that can be used to solve the structure. Other TPR proteins such as $BcsC^{TPR\ 1\text{-}6}$ (90) and AlgK (74) have also been solved by this method, thereby, offering further hope for success in this area.

Possessing solved crystal structures combined with SAXS data of overlapping constructs that would span the entire BcsC TPR domain has begun to answer questions regarding the overall orientation within the periplasm and may also provide evidence of conserved residues used in protein-protein interactions related to the overall function of the protein. Future evidence presented in other solved structures may confirm the theory that numerous TPR folds present regions important in ligand binding and protein-protein interactions to generate a fully functioning biosynthetic complex. Examples of numerous TPR structures exist that illustrate the use of folds as platforms or scaffold for protein interactions that can bind through hydrophobic and hydrophilic character and others through stronger forces, such as hydrogen bonding (87, 88). In the case of BcsC, it seems logical that some type of strong yet reversible interactions with BcsZ occur through binding on the convex surface of the BcsC, which facilitates hydrolysis and release of cellulose chains travelling through BcsCs concave surface. However, this is only speculation based on current knowledge, and a resolved structure of overlapping BcsC protein constructs may elucidate facts regarding the specific structure and function of this process.

6. Integrative Nature of This Research

The discipline of biology encompasses many fields, and this is also true for the sub-discipline of protein crystallography. The research conducted throughout this project provided a substantial framework for learning about dozens of specialized fields. For example, the initial identification of bacterial proteins requires the knowledge to search and navigate seemingly complicated computer programs that utilized complex mathematical algorithms to make informed decisions on protein stability, construct generation and purification optimization long before the initial wet lab research even began. Perhaps even prior to that, was the idea that a specific species of bacteria was creating a niche for study through its ubiquitous behaviour in the environment or in clinical settings. One such group of bacteria, the Enterobacteriaceae, is home to a large family of commensal and pathogenic species that include *Escherichia coli, Shigella, Salmonella*, and countless others that can exist within a community of microbial species, termed a biofilm. Billions of dollars have been funneled into the study of biofilms and into bacterial cellulose-containing biofilms in particular. The pure polymer form of bacterial cellulose has found relevance in a new and exciting field of nanocellulose technology that has applications in biotechnological, medical, pharmaceutical, and food industries. These developments/applications are also rivalled by the interest of researchers in the role cellulose plays in the survival of bacterial biofilms through providing increased resistance to anti-microbial agents and protection from the immune system. The scope of this research focused on the cultivation of lab strain species of bacteria that were used for gene expression in the overproduction of protein for the

purpose of downstream applications that were not limited to, but highlighted X-ray crystallography and computer algorithms to solve protein structures. Throughout this process the disciplines of microbiology, structural biology, the aforementioned biochemistry, analytical chemistry, physics, math, computer programs and bioinformatics were all combined to generate consistent and reliable results. Overall, the nature of the present research is widely integrative, and to specialize in this field, means mastering many aspects from numerous others.

7. REFERENCES

1. World Health Organization (2015) *World Health Statistics*

2. Scharff, R. L. *Economic Burden from Health Losses Due to Foodborne Illness in the United States*, 10.4315/0362-028X.JFP-11-058

3. United States Department of Agriculture Economic Research Service (2015) *Cost Estimates of Foodborne Illnesses*

4. Di Domenico, E. G., Cavallo, I., Pontone, M., Toma, L., and Ensoli, F. (2017) Biofilm Producing Salmonella Typhi: Chronic Colonization and Development of Gallbladder Cancer. *Int. J. Mol. Sci.* **18**, 1887

5. Monteiro, C., Saxena, I., Wang, X., Kader, A., Bokranz, W., Simm, R., Nobles, D., Chromek, M., Brauner, A., Brown, R. M., and Römling, U. (2009) Characterization of cellulose production in Escherichia coli Nissle 1917 and its biological consequences. *Environ. Microbiol.* **11**, 1105–1116

6. Rangel, J. M., Sparling, P. H., Crowe, C., Griffin, P. M., and Swerdlow, D. L. (2005) Epidemiology of *Escherichia coli* O157:H7 Outbreaks, United States, 1982–2002. *Emerg. Infect. Dis.* **11**, 603–609

7. Salvadori, M. I., Sontrop, J. M., Garg, A. X., Moist, L. M., Suri, R. S., and Clark, W. F. (2009) Factors that led to the Walkerton tragedy. *Kidney Int.* **75**, S33–S34

8. Heiman, K. E., Mody, R. K., Johnson, S. D., Griffin, P. M., and Hannah Gould, L. (2015) Escherichia coli O157 Outbreaks in the United States, 2003–2012. *Emerg. Infect. Dis.* **21**, 1293–1301

9. Centers for Disease Control and Prevention (CDC) (2016) Multistate outbreak of shiga toxin-producing Escherichia coli infections linked to flour. [online] https://www.cdc.gov/ecoli/2016/o157h7-09-16/index.html (Accessed March 16, 2017)

10. Canadian Food Inspection Agency (2017) Updated food recall warning – various brands of pie and tart shells recalled due to E. coli O121. https://www.newswire.ca/news-releases/updated-food-recall-warning---various-brands-of-pie-and-tart-shells-recalled-due-to-e-coli-o121-626054513.html. [online] http://www.inspection.gc.ca/about-the-cfia/newsroom/food-recall-warnings/complete-listing/2017-05-26/eng/1495854753520/1495854756704 (Accessed September 5, 2018)

11. Centers for Disease Control and Prevention (2016) Multistate Outbreak of Multidrug-Resistant Salmonella Heidelberg Infections Linked to Contact with Dairy Bull Calves. [online] https://www.cdc.gov/salmonella/heidelberg-11-16/index.html (Accessed March 16, 2017)

12. Costerton, W., Cheng, K.-L., Geesey, G. G., Ladd, T. L., Dasgupta, M., and Marrie, T. I. (1987) Bacterial Biofilms in Nature and Disease. *Ann. Rev. Microbiol*. **41**, 435–64

13. Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R., and Lappin-Scott, H. M. (1995) Microbial Biofilms. *Annu. Rev. Microbiol.* **49**, 711–745

14. Prakash, B., Veeregowda, B. M., and Krishnappa, G. (2003) <Biofilms- A survival

strategy of bacteria.pdf>. *Curr. Sci.*

15. Costerton, J. W., Stewart, P. S., and Greenberg, E. P. (1999) Bacterial biofilms: A common cause of persistent infections. *Science (80-.).* **284**, 1318–1322

16. Fux, C. A., Costerton, J. W., Stewart, P. S., and Stoodley, P. (2005) Survival strategies of infectious biofilms. *Trends Microbiol.* **13**, 34–40

17. Stewart, P. S., and William Costerton, J. (2001) Antibiotic resistance of bacteria in biofilms. *Lancet*. **358**, 135–138

18. Donlan, R. M. (2001) Biofilms and device-associated infections. in *Emerging Infectious Diseases*, pp. 277–281, Centers for Disease Control and Prevention, **7**, 277–281

19. Lewis, K. (2001) Riddle of biofilm resistance. *Antimicrob. Agents Chemother.* **45**, 999–1007

20. Gualdi, L., Tagliabue, L., Bertagnoli, S., Ierano, T., De Castro, C., and Landini, P. (2008) Cellulose modulates biofilm formation by counteracting curli-mediated colonization of solid surfaces in Escherichia coli. *Microbiology*. **154**, 2017–2024

21. Wei, Q., and Ma, L. Z. (2013) Biofilm matrix and its regulation in Pseudomonas aeruginosa. *Int. J. Mol. Sci.* **14**, 20983–21005

22. Cergole-novella, M. C., Pignatari, A. C. C., and Guth, B. E. C. (2015) Adhesion , biofilm and genotypic characteristics of antimicrobial resistant Escherichia coli isolates. *Brazilian J. Microbiol.* **171**, 167–171

23. Høiby, N., Bjarnsholt, T., Givskov, M., Molin, S., and Ciofu, O. (2010) Antibiotic

resistance of bacterial biofilms. *Int. J. Antimicrob. Agents*. **35**, 322–332

24. Hengge, R. (2009) Principles of c-di-GMP signalling in bacteria. *Nat. Rev. Microbiol.* **7**, 263–273

25. Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M., and Jenal, U. (2006) Allosteric control of cyclic di-GMP signaling. *J. Biol. Chem.* **281**, 32015–24

26. Tamayo, R., Pratt, J. T., and Camilli, A. (2007) Roles of Cyclic Diguanylate in the Regulation of Bacterial Pathogenesis. *Annu. Rev. Microbiol.* **61**, 131–148

27. Ryan, R. P., Fouhy, Y., Lucey, J. F., Crossman, L. C., Spiro, S., He, Y.-W., Zhang, L.-H., Heeb, S., Cámara, M., Williams, P., and Dow, J. M. (2006) Cell-cell signaling in Xanthomonas campestris involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6712–7

28. Ryjenkov, D. A., Tarutina, M., Moskvin, O. V, and Gomelsky, M. (2005) Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *J. Bacteriol.* **187**, 1792–8

29. Schmidt, A. J., Ryjenkov, D. A., and Gomelsky, M. (2005) The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J. Bacteriol.* **187**, 4774–81

30. Simm, R., Morr, M., Kader, A., Nimtz, M., and Römling, U. (2004) GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol. Microbiol.* **53**, 1123–1134

31. Ryjenkov, D. A., Simm, R., Römling, U., and Gomelsky, M. (2006) The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. *J. Biol. Chem.* **281**, 30310–4

32. Romling, U., Galperin, M. Y., and Gomelsky, M. (2013) Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. *Microbiol. Mol. Biol. Rev.* **77**, 1–52

33. Zorraquino, V., García, B., Latasa, C., Echeverz, M., Toledo-Arana, A., Valle, J., Lasa, I., and Solano, C. (2013) Coordinated cyclic-di-GMP repression of Salmonella motility through YcgR and cellulose. *J. Bacteriol.* **195**, 417–28

34. Yang, F., Tian, F., Sun, L., Chen, H., Wu, M., Yang, C.-H., and He, C. (2012) A Novel Two-Component System PdeK/PdeR Regulates c-di-GMP Turnover and Virulence of Xanthomonas oryzae pv. oryzae. */ 1361 MPMI*. 10.1094/MPMI-01-12-0014-R

35. Miller, S. I., Kukral, A. M., and Mekalanos, J. J. (1989) A two-component regulatory system (phoP phoQ) controls Salmonella typhimurium virulence. *Proc. Natl. Acad. Sci.* **86**, 5054–5058

36. Xu, L., Venkataramani, P., Ding, Y., Liu, Y., Deng, Y., Yong, G. L., Xin, L., Ye, R., Zhang, L., Yang, L., and Liang, Z.-X. (2016) A Cyclic di-GMP-binding Adaptor Protein Interacts with Histidine Kinase to Regulate Two-component Signaling. *J. Biol. Chem.* **291**, 16112–23

37. Tal, R., Wong, H. C., Calhoon, R., Gelfand, D., Fear, A. L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., Cohen, A., Sapir, S., Ohana, P., and Benziman, M. (1998) Three cdg operons control cellular turnover of cyclic di-GMP in Acetobacter

xylinum: Genetic organization and occurrence of conserved domains in isoenzymes. *J. Bacteriol.* **180**, 4416–4425

38. Jenal, U., and Malone, J. (2006) Mechanisms of Cyclic-di-GMP Signaling in Bacteria Key Words. *Annu. Rev. Genet*. **40**, 385–407

39. Petrova, O. E., and Sauer, K. (2016) Escaping the biofilm in more than one way: desorption, detachment or dispersion. *Curr. Opin. Microbiol.* **30**, 67–78

40. Berlanga, M., and Guerrero, R. (2016) Living together in biofilms: The microbial cell factory and its biotechnological implications. *Microb. Cell Fact.* **15**, 165

41. Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., and Kolter, R. (2013) Sticking together: Building a biofilm the Bacillus subtilis way. *Nat. Rev. Microbiol.* **11**, 157–168

42. Costerton, J. W. (1995) Overview of microbial biofilms. *J. Ind. Microbiol.* **15**, 137–140

43. Molin, S., and Tolker-Nielsen, T. (2003) Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Curr. Opin. Biotechnol.* **14**, 255–261

44. Boyd, C. D., and O'Toole, G. A. (2012) Second Messenger Regulation of Biofilm Formation: Breakthroughs in Understanding c-di-GMP Effector Systems. *Annu. Rev. Cell Dev. Biol.* **28**, 439–462

45. Cotter, P. A., and Stibitz, S. (2007) c-di-GMP-mediated regulation of virulence and biofilm formation. *Curr. Opin. Microbiol.* **10**, 17–23

46. de Beer, D., Stoodley, P., Roe, F., and Lewandowski, Z. (1994) Effects of biofilm

structures on oxygen distribution and mass transport. *Biotechnol. Bioeng.* **43**, 1131–1138

47. Karatan, E., and Watnick, P. (2009) Signals, Regulatory Networks, and Materials That Build and Break Bacterial Biofilms. *Microbiol. Mol. Biol. Rev.* **73**, 310–347

48. Lewis, K. (2010) Persister Cells. *Annu. Rev. Microbiol.* **64**, 357–372

49. Lewis, K. (2007) Persister cells, dormancy and infectious disease. *Nat. Rev. Microbiol.* **5**, 48–56

50. Wood, T. K., Knabel, S. J., and Kwan, B. W. (2013) Bacterial persister cell formation and dormancy. *Appl. Environ. Microbiol.* **79**, 7116–21

51. Kaplan, J. B. (2010) Biofilm Dispersal. *J. Dent. Res.* **89**, 205–218

52. Gupta, K., Liao, J., Petrova, O. E., Cherny, K. E., and Sauer, K. (2014) Elevated levels of the second messenger c-di-GMP contribute to antimicrobial resistance of *P seudomonas aeruginosa*. *Mol. Microbiol.* **92**, 488–506

53. Davis, P. B., Drumm, M., and Konstan, M. W. (1996) Cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **154**, 1229–56

54. Ahmed, S. A. S., Dooley, J. S., Deb, S. R., Talukder, D. C., Hossain, S. S., and Anwar, A. T. (2013) Pseudomonas aeruginosa infection in cystic fibrosis patients. *J. Dhaka Med. Coll.* 10.3329/jdmc.v22i1.15689

55. Pendersen, S. S., Hoiby, N., Espersen, F., and Koch, C. (1992) Role of alginate in infection with mucoid Pseudomonas aeruginosa in cystic fibrosis. *Thorax*. **47**, 6–13

56. Yoshiharu Nishiyama, †, Junji Sugiyama, ‡, Henri Chanzy, § and, and Paul Langan*, ∥

(2003) Crystal Structure and Hydrogen Bonding System in Cellulose Iα from Synchrotron X-ray and Neutron Fiber Diffraction. 10.1021/JA037055W

57. Römling, U. (2002) Molecular biology of cellulose production in bacteria. *Res. Microbiol.* **153**, 205–212

58. Ross, P., Mayer, R., and Benziman, M. (1991) Cellulose Biosynthesis and Function in Bacteria. *Microbiol. Rev.* **55**, 35–58

59. Petersen, N., and Gatenholm, P. (2011) Bacterial cellulose-based materials and medical devices: current state and perspectives. *Appl. Microbiol. Biotechnol.* **91**, 1277–1286

60. Czaja, W., Krystynowicz, A., Bielecki, S., and Brown, R. M. (2006) Microbial cellulose - The natural power to heal wounds. *Biomaterials*. **27**, 145–151

61. Ullah, H., Santos, H. A., and Khan, T. (2016) Applications of bacterial cellulose in food, cosmetics and drug delivery. *Cellulose*. **23**, 2291–2314

62. Lin, S. P., Liu, C. Te, Hsu, K. Di, Hung, Y. T., Shih, T. Y., and Cheng, K. C. (2016) Production of bacterial cellulose with various additives in a PCS rotating disk bioreactor and its material property analysis. *Cellulose*. **23**, 367–377

63. Römling, U., and Galperin, M. Y. (2015) Bacterial cellulose biosynthesis: Diversity of operons, subunits, products, and functions. *Trends Microbiol.* **23**, 545–557

64. Sunagawa, N., Tajima, K., Hosoda, M., Kawano, S., Kose, R., Satoh, Y., Yao, M., and Dairi, T. (2012) Cellulose production by Enterobacter sp. CJF-002 and identification of genes for cellulose biosynthesis. *Cellulose*. **19**, 1989–2001

65. Ji, K., Wang, W., Zeng, B., Chen, S., Zhao, Q., Chen, Y., Li, G., and Ma, T. (2016) Bacterial cellulose synthesis mechanism of facultative anaerobe Enterobacter sp. FY-07. *Sci. Rep.* 10.1038/srep21863

66. Zogaj, X., Nimtz, M., Rohde, M., Bokranz, W., and Römling, U. (2001) The multicellular morphotypes of Salmonella typhimurium and Escherichia coli produce cellulose as the second component of the extracellular matrix. *Mol. Microbiol.* **39**, 1452–1463

67. Solano, C., Sesma, B., Alvarez, M., Urdaneta, E., Garcia-Ros, D., Calvo, A., and Gamazo, C. (2001) Virulent strains of Salmonella enteritidis disrupt the epithelial barrier of Caco-2 and HEp-2 cells. *Arch. Microbiol.* **175**, 46–51

68. Solano, C., García, B., Valle, J., Berasain, C., Ghigo, J. M., Gamazo, C., and Lasa, I. (2002) Genetic analysis of Salmonella enteritidis biofilm formation: Critical role of cellulose. *Mol. Microbiol.* **43**, 793–808

69. Whitney, J. C., and Howell, P. L. (2013) Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends Microbiol.* **21**, 63–72

70. Remminghorst, U., and Rehm, B. H. A. (2006) In vitro alginate polymerization and the functional role of Alg8 in alginate production by Pseudomonas aeruginosa. *Appl. Environ. Microbiol.* **72**, 298–305

71. Merighi, M., Lee, V. T., Hyodo, M., Hayakawa, Y., and Lory, S. (2007) The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in Pseudomonas aeruginosa. *Mol. Microbiol.* **65**, 876–895

72. Jain, S., Franklin, M. J., Ertesvåg, H., Valla, S., and Ohman, D. E. (2003) The dual roles of AlgG in C-5-epimerization and secretion of alginate polymers in Pseudomonas aeruginosa. *Mol. Microbiol.* **47**, 1123–1133

73. Weadge, J. T., Yip, P. P., Robinson, H., Arnett, K., Tipton, P. A., and Howell, P. L. (2010) Expression, purification, crystallization and preliminary X-ray analysis of Pseudomonas aeruginosa AlgX. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **66**, 588–91

74. Keiski, C. L., Harwich, M., Jain, S., Neculai, A. M., Yip, P., Robinson, H., Whitney, J. C., Riley, L., Burrows, L. L., Ohman, D. E., and Howell, P. L. (2010) AlgK Is a TPR-Containing Protein and the Periplasmic Component of a Novel Exopolysaccharide Secretin. *Structure*. **18**, 265–273

75. Whitney, J. C., Hay, I. D., Li, C., Eckford, P. D. W., Robinson, H., Amaya, M. F., Wood, L. F., Ohman, D. E., Bear, C. E., Rehm, B. H., and Howell, P. L. (2011) Structural basis for alginate secretion across the bacterial outer membrane. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13083–8

76. Omadjela, O., Narahari, A., Strumillo, J., Melida, H., Mazur, O., Bulone, V., and Zimmer, J. (2013) BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. *Proc. Natl. Acad. Sci.* **110**, 17856–17861

77. Morgan, J. L. W., Strumillo, J., and Zimmer, J. (2012) Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature*. **493**, 181–186

78. McNamara, J. T., Morgan, J. L. W., and Zimmer, J. (2015) A Molecular Description of

Cellulose Biosynthesis. *Annu. Rev. Biochem.* **84**, 895–921

79. Mazur, O., and Zimmer, J. (2011) Apo- and cellopentaose-bound structures of the bacterial cellulose synthase subunit BcsZ. *J. Biol. Chem.* **286**, 17601–6

80. Hirano, T., Kinoshita, N., Morikawa, K., and Yanagida, M. (1990) Snap helix with knob and hole: Essential repeats in S. pombe nuclear protein nuc2 +. *Cell*. **60**, 319–328

81. Blatch, G. L., and Lässle, M. (1999) The tetratricopeptide repeat: A structural motif mediating protein-protein interactions. *BioEssays*. **21**, 932–939

82. D'Andrea, L. D., and Regan, L. (2003) TPR proteins: the versatile helix. *Trends Biochem. Sci.* **28**, 655–662

83. Zeytuni, N., and Zarivach, R. (2012) The TPR Motif as a Protein Interaction Module - A Discussion of Structure and Function. in *Protein Interactions*, InTech, 10.5772/37566

84. Jeong, H., Sim, H. J., Song, E. K., Lee, H., Ha, S. C., Jun, Y., Park, T. J., and Lee, C. (2016) Crystal structure of SEL1L: Insight into the roles of SLR motifs in ERAD pathway. *Sci. Rep.* **6**, 20261

85. Sikorski, R. S., Boguski, M. S., Goebl, M., and Hieter, P. (1990) A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell*. **60**, 307–317

86. Sawyer, N., Chen, J., and Regan, L. (2013) All repeats are not equal: a module-based approach to guide repeat protein design. *J. Mol. Biol.* **425**, 1826–1838

87. Zeytuni, N., Ozyamak, E., Ben-Harush, K., Davidov, G., Levin, M., Gat, Y., Moyal, T.,

Brik, A., Komeili, A., and Zarivach, R. (2011) Self-recognition mechanism of MamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. *Proc. Natl. Acad. Sci.* **108**, E480–E487

88. Zeytuni, N., and Zarivach, R. (2012) Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module. *Structure*. **20**, 397–405

89. Wong, H. C., Fear, A. L., Calhoont, R. D., Eichinger, G. H., Mayer, R., Amikam, D., Benziman, M., Gelfand, D. H., Meade, J. H., Emerick, A. W., Bruner, R., Ben-Bassat, A., and Tal, R. (1990) Genetic organization of the cellulose synthase operon in Acetobacter xylinum. *Genetics*. **87**, 8130–8134

90. Nojima, S., Fujishima, A., Kato, K., Ohuchi, K., Shimizu, N., Yonezawa, K., Tajima, K., and Yao, M. (2017) Crystal structure of the flexible tandem repeat domain of bacterial cellulose synthesis subunit C. *Sci. Rep.* **7**, 13018

91. Vuong, C., Kocianova, S., Voyich, J. M., Yao, Y., Fischer, E. R., DeLeo, F. R., and Otto, M. (2004) A crucial role for exopolysaccharide modification in bacterial biofilm formation, immune evasion, and virulence. *J. Biol. Chem.* **279**, 54881–6

92. Centers for Disease Control and Prevention (2017) Reports of Selected E. coli Outbreak Investigations. [online] https://www.cdc.gov/ecoli/outbreaks.html (Accessed March 22, 2017)

93. Multistate Outbreak of E. coli O157:H7 Infections Linked to Romaine Lettuce (Final Update) | Investigation Notice: Multistate Outbreak of E. coli O157:H7 Infections April 2018 | E. coli | CDC *2018*. [online] https://www.cdc.gov/ecoli/2018/o157h7-04-

18/index.html (Accessed September 6, 2018)

94. Corcoran, M., Morris, D., De Lappe, N., O'Connor, J., Lalor, P., Dockery, P., and Cormican, M. (2014) Commonly used disinfectants fail to eradicate Salmonella enterica biofilms from food contact surface materials. *Appl. Environ. Microbiol.* **80**, 1507–14

95. CDC (2016) Outbreaks Involving Salmonella | CDC. *CDC*. [online] https://www.cdc.gov/salmonella/outbreaks.html (Accessed March 22, 2017)

96. Centers for Disease Control and Prevention (2015) CDC Current Outbreak List. [online] https://www.cdc.gov/outbreaks/index.html (Accessed September 6, 2018)

97. Monds, R. D., and O'Toole, G. A. (2009) The developmental model of microbial biofilms: ten years of a paradigm up for review. *Trends Microbiol.* **17**, 73–87

98. Oliveira, N. M., Martinez-Garcia, E., Xavier, J., Durham, W. M., Kolter, R., Kim, W., and Foster, K. R. (2015) Biofilm formation as a response to ecological competition. *PLoS Biol.* **13**, e1002191

99. Beloin, C., Roux, A., and Ghigo, J. M. (2008) Escherichia coli biofilms. *Curr. Top. Microbiol. Immunol.* **322**, 249–89

100. Corcoran, M., Morris, D., De Lappe, N., O'Connor, J., Lalor, P., Dockery, P., and Cormican, M. (2014) Commonly used disinfectants fail to eradicate Salmonella enterica biofilms from food contact surface materials. *Appl. Environ. Microbiol.* **80**, 1507–1514

101. Wang, R., Bono, J. L., Kalchayanand, N., Shackelford, S., and Harhay, D. M. (2012) Biofilm formation by Shiga toxin-producing Escherichia coli O157:H7 and Non-O157 strains and their tolerance to sanitizers commonly used in the food processing

environment. *J. Food Prot.* **75**, 1418–28

102. Wang, Y., Moradali, M. F., Goudarztalejerdi, A., Sims, I. M., and Rehm, B. H. A. (2016) Biological function of a polysaccharide degrading enzyme in the periplasm. *Sci. Rep.* **6**, 31249

103. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook*, pp. 571–607, 10.1385/1592598900

104. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858

105. DeLano, W. L. (2002) The PyMOL Molecular Graphics System. *Schrödinger LLC wwwpymolorg*. **Version 1.**, http://www.pymol.org

106. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402

107. Bhagwat, M., and Aravind, L. (2007) PSI-BLAST tutorial. *Methods Mol. Biol.* **395**, 177–186

108. Sievers, F., and Higgins, D. G. (2014) Clustal Omega. *Curr. Protoc. Bioinforma.* **2014**, 3.13.1-3.13.16

109. Atkins, J., Boateng, S., Sorensen, T., and McGuffin, L. (2015) Disorder Prediction Methods, Their Applicability to Different Protein Targets and Their Usefulness for Guiding Experimental Studies. *Int. J. Mol. Sci.* **16**, 19040–19054

110. Yoshida, N., and Sato, M. (2009) Plasmid uptake by bacteria: A comparison of methods and efficiencies. *Appl. Microbiol. Biotechnol.* **83**, 791–798

111. Bondos, S. E., and Bicknell, A. (2003) Detection and prevention of protein aggregation before, during, and after purification. *Anal. Biochem.* **316**, 223–231

112. Arzenšek, D., Podgornik, R., and Kuzman, D. (2010) Dynamic light scattering and application to proteins in solutions. *Semin. Dep. Physics, Univ. Ljubljana*

113. Stura, E. A., and Wilson, I. A. (1991) *Applications of the streak seeding technique in protein crystallization*

114. Luft, J. R., Collins, R. J., Fehrman, N. A., Lauricella, A. M., Veatch, C. K., and DeTitta, G. T. (2003) A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J. Struct. Biol.* **142**, 170–179

115. Haupert, L. M., and Simpson, G. J. (2011) Screening of protein crystallization trials by second order nonlinear optical imaging of chiral crystals (SONICC). *Methods*. **55**, 379–86

116. Madden, J. T., DeWalt, E. L., and Simpson, G. J. (2011) Two-photon excited UV fluorescence for protein crystal detection. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 839–46

117. Fodje, M., Janzen, K., Berg, R., Black, G., Labiuk, S., Gorin, J., and Grochulski, P. (2012) MxDC and MxLIVE: Software for data acquisition, information management and remote access to macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **19**, 274–280

118. Bunkóczi, G., and Read, R. J. (2011) Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 303–312

119. Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221

120. Bunkóczi, G., Echols, N., McCoy, A. J., Oeffner, R. D., Adams, P. D., and Read, R. J. (2013) Phaser.MRage: Automated molecular replacement. in *Acta Crystallographica Section D: Biological Crystallography*, pp. 2276–2286, International Union of Crystallography, **69**, 2276–2286

121. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242

122. Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2007) BALBES: A molecular-replacement pipeline. in *Acta Crystallographica Section D: Biological Crystallography*, pp. 125–132, International Union of Crystallography, **64**, 125–132

123. Keegan, R. M., and Winn, M. D. (2007) Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 447–457

124. Lebedev, A., and Vagin, A. (2015) MoRDa, an automatic molecular replacement pipeline. *Acta Crystallogr. Sect. A Found. Adv.* **71**, 4

125. Classen, S., Hura, G. L., Holton, J. M., Rambo, R. P., Rodic, I., McGuire, P. J., Dyer, K., Hammel, M., Meigs, G., Frankel, K. A., and Tainer, J. A. (2013) Implementation and performance of SIBYLS: A dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Crystallogr.* **46**, 1–13

126. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., Tsutakawa, S. E., Jenney, F. E., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods*. **6**, 606–612

127. Dyer, K. N., Hammel, M., Rambo, R. P., Tsutakawa, S. E., Rodic, I., Classen, S., Tainer, J. A., and Hura, G. L. (2014) High-throughput SAXS for the characterization of biomolecules in solution: A practical approach. *Methods Mol. Biol.* **1091**, 245–258

128. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285

129. Shin, D. (2017) SAXS FrameSlice. [online] http://sibyls.als.lbl.gov/ran (Accessed September 14, 2018)

130. Förster, S., Apostol, L., and Bras, W. (2010) Scatter: Software for the analysis of nano- and mesoscale small-angle scattering. *J. Appl. Crystallogr.* **43**, 639–646

131. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503

132. Konarev, P. V, Volkov, V. V, Sokolova, A. V, Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS : a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282

133. Svergun, D. I., Petoukhov, M. V, and Koch, M. H. J. (2001) Determination of domain structure of proteins from x-ray solution scattering. *Biophys. J.* **80**, 2946–2953

134. Franke, D., Petoukhov, M. V, Konarev, P. V, Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M., and Svergun, D. I. (2017) ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225

135. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864

136. Panjkovich, A., and Svergun, D. I. (2016) SASpy: A PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data. *Bioinformatics*. **32**, 2062–2064

137. Scapin, G. (2013) Molecular replacement then and now. *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 2266–75

138. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501

139. Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F.,

Söding, J., Lupas, A. N., and Alva, V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243

140. Structural Genomics Consortium, S. G., China Structural Genomics Consortium, A. et F. des M., Northeast Structural Genomics Consortium, B. S. G., Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., Dhe-Paganon, S., Park, H., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S.-H., Rao, Z., Shi, Y., Terwilliger, T. C., Kim, C.-Y., Hung, L.-W., Waldo, G. S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J. L., Stevens, R. C., Lesley, S. A., Wilson, I. A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M. I., Eschenfeldt, W. H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S. K., Emtage, J. S., Sauder, J. M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S. C., Bonanno, J. B., Fiser, A., Swaminathan, S., Studier, F. W., Chance, M. R., Sali, A., Acton, T. B., Xiao, R., Zhao, L., Ma, L. C., Hunt, J. F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C. K., Wang, D., Wang, H., Jiang, M., Montelione, G. T., Stuart, D. I., Owens, R. J., Daenke, S., Schütz, A., Heinemann, U., Yokoyama, S., Büssow, K., and Gunsalus, K. C. (2008) Protein production and purification. *Nat. Methods*. **5**, 135–46

141. Pike, A. C. W., Garman, E. F., Krojer, T., Von Delft, F., and Carpenter, E. P. (2016) An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 303–318

142. Skou, S., Gillilan, R. E., and Ando, N. (2014) Synchrotron-based small-angle X-ray

scattering of proteins in solution. *Nat. Protoc.* **9**, 1727–1739

143. Jacques, D. A., and Trewhella, J. (2010) Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls. *Protein Sci.* **19**, 642–657

144. Strub, M. P., Hoh, F., Sanchez, J. F., Strub, J. M., Böck, A., Aumelas, A., and Dumas, C. (2003) Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure*. **11**, 1359–1367

145. Kadokura, H., Katzen, F., and Beckwith, J. (2003) Protein Disulfide Bond Formation in Prokaryotes. *Annu. Rev. Biochem.* **72**, 111–135

146. Weyer, F. A., Gumiero, A., Lapouge, K., Bange, G., Kopp, J., and Sinning, I. (2017) Structural basis of HypK regulating N-terminal acetylation by the NatA complex. *Nat Commun*. **8**, 15726–15726

147. Neubauer, J. L., Pham, T., Immormino, R. M., Dollins, D. E., Endo-Streeter, S. T., Li, S., Pemble IV, C. W., and York, J. D. The Protein Complex NatA Binds Inositol Hexakisphosphate and Exhibits Conformational Flexibility. *TO BE Publ.* 10.2210/PDB4HNX/PDB

148. Jínek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A., and Conti, E. (2004) The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. *Nat. Struct. Mol. Biol.* **11**, 1001–1007

149. Halbach, F., Reichelt, P., Rode, M., and Conti, E. (2013) The Yeast Ski Complex: Crystal Structure and RNA Channeling to the Exosome Complex. *Cell*. **154**, 814

150. Taschner, M., Kotsis, F., Braeuer, P., Kuehn, E. W., and Lorentzen, E. (2014) Crystal

Structures of Ift70/52 and Ift52/46 Provide Insight Into Intraflagellar Transport B Core Complex Assembly. *J.Cell Biol.* **207**, 269

151. Hay, I. D., Schmidt, O., Filitcheva, J., and Rehm, B. H. A. (2012) Identification of a periplasmic AlgK-AlgX-MucD multiprotein complex in Pseudomonas aeruginosa involved in biosynthesis and regulation of alginate. *Appl. Microbiol. Biotechnol.* **93**, 215–227

152. Fata Moradali, M., Donati, I., Sims, I. M., Ghods, S., and Rehm, B. H. A. (2015) Alginate polymerization and modification are linked in pseudomonas aeruginosa. *MBio*. **6**, 1–17

153. Ishida, T., and Kinoshita, K. (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, W460–W464

154. McPherson, A., and Gavira, J. A. (2014) Introduction to protein crystallization. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70**, 2–20

155. Atkins, J. D., Boateng, S. Y., Sorensen, T., and McGuffin, L. J. (2015) Disorder prediction methods, their applicability to different protein targets and their usefulness for guiding experimental studies. *Int. J. Mol. Sci.* **16**, 19040–19054

156. A. Keith Dunker, *, ‡, Celeste J. Brown, ‡, J. David Lawson, ‡, Lilia M. Iakoucheva, ‡ and, and Obradović§, Z. (2002) Intrinsic Disorder and Protein Function†. *Biochemistry*. **41**, 6573–6582

157. Huang, Y. J., Acton, T. B., and Montelione, G. T. (2014) DisMeta: A Meta Server for Construct Design and Optimization, pp. 3–16, 10.1007/978-1-62703-691-7_1

158. Ishida, T., and Kinoshita, K. (2008) Prediction of disordered regions in proteins based on

the meta approach. *Bioinformatics*. **24**, 1344–1348

159. Huang, Y. J., Acton, T. B., and Montelione, G. T. (2014) DisMeta: a meta server for construct design and optimization. *Methods Mol. Biol.* **1091**, 3–16

160. Galloway, C. A., Sowden, M. P., and Smith, H. C. (2003) Increasing the yield of soluble recombinant protein expressed in E. coli by induction during late log phase. *Biotechniques*. **34**, 524–6, 528, 530

161. Waugh, D. S. (2005) Making the most of affinity tags. *Trends Biotechnol.* **23**, 316–320

162. Low, K. E., and Howell, P. L. (2018) Gram-negative synthase-dependent exopolysaccharide biosynthetic machines. *Curr. Opin. Struct. Biol.* **53**, 32–44

163. Silhavy, T. J., Kahne, D., and Walker, S. (2010) The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2**, a000414

164. Drenkard, E. (2003) Antimicrobial resistance of Pseudomonas aeruginosa biofilms. *Microbes Infect.* **5**, 1213–1219

165. Mongkolrob, R., Taweechaisupapong, S., and Tungpradabkul, S. (2015) Correlation between biofilm production, antibiotic susceptibility and exopolysaccharide composition in Burkholderia pseudomallei bpsI, ppk, and rpoS mutant strains. *Microbiol. Immunol.* **59**, 653–663

166. Ullah, H., Santos, H. A., and Khan, T. (2016) Applications of bacterial cellulose in food, cosmetics and drug delivery. *Cellulose*. **23**, 2291–2314

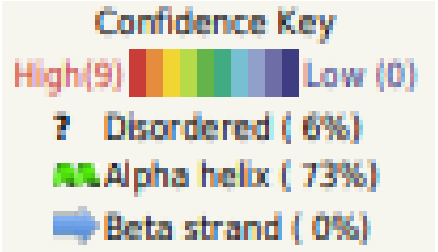
167. Derewenda, Z. S. (2004) The use of recombinant methods and molecular engineering in

protein crystallization. *Methods*. **34**, 354–363

8. APPENDIX


Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence
MAPTAQQQLLEQVRLGEATHREDLVQQSLYRLELIDPNNPDVVAARFRSLLRQGDIDGAQ
KQLDRLSQLAPSSNAYKSSRTTMLLSTPDGRQALQQARLQATTGHAEEAVASYNKLFNGA
PPEGDIAVEYWSTVAKIPARRGEAINQLKRINADAPGNTGLQNNLALLLFSSDRRDEGFA
VLEQMAKSNAGREGASKIWYGQIKDMPVSDASVSALKKYLSIFSDGDSVAAAQSQLAEQQ
KQLADPAFRARAQGLAAVDSGMAGKAIPELQQAVRANPKDSEALGALGQAYSQKGDRANA
VANLEKALALDPHSSNNDKWNSLLKVNRYWLAIQQGDAALKANNPDRAERLFQQARNVDN
TDSYAVLGLGDVAMARKDYPAAERYYQQTLRMDSGNTNAVRGLANIYRQQSPEKAEAFIA


Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence


Confidence Key
High(9)
Low (0)
? Disordered (6%)
Alpha helix (73%)
Beta strand (0%)

Escherichia coli (strain K12) Cellulose Synthase Operon Protein C

Signal Peptide aa 1-23: MRKFTLNIFTLSLGLAVMPMVEA

Full Length BcsC Construct aa 24-1157:

APTAQQQLLEQVRLGEATHREDLVQQSLYRLELIDPNNPDVVAARFRSLLRQGDIDGAQK
QLDRLSQLAPSSNAYKSSRTTMLLSTPDGRQALQQARLQATTGHAEEAVASYNKLFNGAP
PEGDIAVEYWSTVAKIPARRGEAINQLKRINADAPGNTGLQNNLALLLFSSDRRDEGFAV
LEQMAKSNAGREGASKIWYGQIKDMPVSDASVSALKKYLSIFSDGDSVAAAQSQLAEQQK
QLADPAFRARAQGLAAVDSGMAGKAIPELQQAVRANPKDSEALGALGQAYSQKGDRANAV
ANLEKALALDPHSSNNDKWNSLLKVNRYWLAIQQGDAALKANNPDRAERLFQQARNVDNT
DSYAVLGLGDVAMARKDYPAAERYYQQTLRMDSGNTNAVRGLANIYRQQSPEKAEAFIAS
LSASQRRSIDDIERSLQNDRLAQQAEALENQGKWAQAAALQRQRLALDPGSVWITYRLSQ
DLWQAGQRSQADTLMRNLAQQKSNDPEQVYAYGLYLSGHDQDRAALAHINSLPRAQWNSN
IQELVNRLQSDQVLETANRLRESGKEAEAEAMLRQQPPSTRIDLTLADWAQQRRDYTAAR
AAYQNVLTREPANADAILGLTEVDIAAGDKAAARSQLAKLPATDNASLNTQRRVALAQAQ
LGDTAAAQRTFNKLIPQAKSQPPSMESAMVLRDGAKFEAQAGDPTQALETYKDAMVASGV
TTTRPQDNDTFTRLTRNDEKDDWLKRGVRSDAADLYRQQDLNVTLEHDYWGSSGTGGYSD
LKAHTTMLQVDAPYSDGRMFFRSDFVNMNVGSFSTNADGKWDDNWGTCTLQDCSGNRSQS
DSGASVAVGWRNDVWSWDIGTTPMGFNVVDVVGGISYSDDIGPLGYTVNAHRRPISSSLL
AFGGQKDSPSNTGKKWGGVRADGVGLSLSYDKGEANGVWASLSGDQLTGKNVEDNWRVRW
MTGYYYKVINQNNRRVTIGLNNMIWHYDKDLSGYSLGQGGYYSPQEYLSFAIPVMWRERT
ENWSWELGASGSWSHSRTKTMPRYPLMNLIPTDWQEEAARQSNDGGSSQGFGYTARALLE
RRVTSNWFVGTAIDIQQAKDYAPSHFLLYVRYSAAGWQGDMDLPPQPLIPYADW